www.freeservicemanuals.info 11/14/2023

PHILIPS TECHNISCH TIJDSCHRIFT

JAARGANG 32, 1971, no. 10/11/12

Op 15 juni '72 heeft prof. dr. H. B. G. Casimir, tot dan lid van de Raad van Bestuur van het Philips-concern en als zodanig leider van de Philips-research, zijn functie neergelegd. Zoals reeds eerder op deze plaats is gezegd, laat de aard van ons tijdschrift het niet toe een beschouwing aan een dergelijk gebeuren te wijden. Wel echter heeft de redactie gemeend prof. Casimir — die vele jaren de directe chef van de redactie is geweest en die altijd een grote belangstelling voor Philips technisch Tijdschrift heeft getoond — als lezer van het tijdschrift een buiten zijn medeweten voorbereid nummer te moeten aanbieden, dat enerzijds is samengesteld met als richtsnoer zijn belangstellingssfeer maar dat ook een zekere weerspiegeling vormt van aard en omvang van zijn functie. Een aantal van de in dit nummer opgenomen artikelen behoort daardoor tot een type dat in ons tijdschrift meer uitzondering dan regel is, maar daar staan andere tegenover die er juist bij uitstek in passen. Naast een drietal bijdragen van medewerkers van het Natuurkundig Laboratorium te Eindhoven zijn er bijdragen uit ieder der buitenlandse researchlaboratoria: Philips Forschungslaboratorium Hamburg, Philips Forschungslaboratorium Aachen, de Laboratoires d'Electronique et de Physique Appliquée te Limeil-Brévannes, het MBLE Laboratoire de Recherches te Brussel, Mullard Research Laboratories te Redhill. Bovendien is er een bijdrage van Philips Laboratories te Briarcliff Manor, New York. Afgezien van een kleine redactionele escapade, opent het nummer met een beschouwing van prof. Casimir zelf; de redactie heeft gemeend er in dit geval goed aan te doen de lezer naar aanleiding van wiens aftreden dit nummer is samengesteld, zelf aan zijn afscheidscadeau te laten „meebetalen".

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

De wetenschap en haar beoefenaars in vroeger eeuwen

De onder deze titel bijeengebrachte stukken zijn citaten uit het vermaarde boek Gulliver's Travels van Jonathan Swift (London 1726; Nederlandse vertaling van mr. S. Davids, uitgeverij Paris, Amsterdam 1957). Ze zijn alle ontleend aan het derde deel, en wel aan de hoofdstukken over Gulliver's verblijf op de eilanden Laputa en Balnibarbi.

Het eiland Laputa zweeft in de omgeving van Balnibarbi in de lucht dank zij de kracht van een reusachtige, in zijn zwaartepunt door een diamanten as ondersteunde magneet, waarmee men het eiland kan laten stijgen, dalen en zich in horizontale richting verplaatsen. De stand van de magneet wordt door sterrenkundigen ingesteld, uitgaande van de richting waarin de vorst het eiland verplaatst wil zien. De mannelijke bewoners van Laputa, de vorst inbegrepen, zijn zeer bedreven in wiskunde en muziek, maar interesseren zich nauwelijks voor iets anders.

Het eiland Balnibarbi — een gewoon eiland — staat eveneens onder het bestuur van de vorst van Laputa; de hoofdstad ervan is de stad Lagado, alwaar een Grote Academie is gevestigd.

Gulliver over Laputa en zijn bewoners...

Mijn kennis van de wiskunde kwam mij goed te stade bij het aanleren van hun spreektrant, die sterk samenhangt met die wetenschap en met de muziek, waarin ik ook niet onbedreven ben. Hun gedachten houden zich voortdurend bezig met lijnen en figuren.

* * *

Hun huizen zijn heel slecht gebouwd, de muren geren, in geen enkele kamer is er ook maar één rechte hoek; en dit gebrek komt voort uit hun minachting voor practische meetkunde, waar zij op neerzien als op iets banaals en mechanisch. Hun opdrachten zijn te ingewikkeld voor de geestvermogens van hun werklieden, hetgeen eindeloze misverstanden teweegbrengt. En al zijn ze heel bekwaam in het hanteren van liniaal, potlood en passer op een vel papier, in de gewone dingen en gedragingen des levens heb ik nooit een schichtiger, stunteliger en onhandiger volk gezien, of een dat zo traag en verward is in zijn begrip van alle zaken behalve wiskunde en muziek. Het zijn heel slechte redeneerders, eeuwig en altijd in de contramine, behalve als ze het toevallig bij het rechte eind hebben, wat maar zelden voorkomt.

* * *

Al kan ik niet beweren dat ik op het eiland slecht behandeld werd, toch moet het mij van het hart, dat ik mij daar te zeer verwaarloosd voelde en zelfs met enige minachting bejegend. Want vorst noch volk leek belang te stellen in enigerlei wetenschap, behalve wiskunde en muziek, waarin ik verre hun mindere was en om die reden zeer weinig in tel.

Daar kwam nog bij, dat ik, na al de merkwaardigheden van het eiland te hebben gezien, bizonder verlangend was het te verlaten, want ik was de mensen daar hartelijk beu. Toegegeven, zij blonken uit in twee wetenschappen waarvoor ik grote achting heb en waarin ik niet onbedreven ben; maar tevens waren zij zo verstrooid en in gepeinzen verdiept, dat het mij niet heugde ooit zulke ongezellige lui te zijn tegengekomen. Ik ging in de twee maanden, dat ik daar toefde, alleen maar om met vrouwen, neringdoenden, wakkerhouders en pages, hetgeen mij allengs tot een algemeen-geminachte figuur maakte; en toch waren dit de enige mensen van wie ik ooit een redelijk antwoord kreeg.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Gulliver bezoekt de Grote Academie van Lagado...

Ik werd zeer vriendelijk door de directeur ontvangen en bezocht de academie vele dagen achtereen. In elke kamer zijn een of meer plannenmakers aan het werk, en ik geloof dat ik in wel vijfhonderd kamers ben geweest.

De eerste man die ik zag was een magere figuur met roetige handen en gelaat, en zijn haren en baard lang, rafelig en op verscheiden plaatsen geschroeid. Zijn kleren, hemd en huid waren alle van dezelfde kleur. Hij was al acht jaar bezig met een plan om zonnestralen uit komkommers te trekken, met de bedoeling ze in hermetisch gesloten flessen te stoppen en ze weer los te laten ter verwarming van de lucht in ruwe en gure zomers. Hij vertelde mij dat hij stellig over nog acht jaar in staat zou zijn, de tuinen van de gouverneur tegen redelijke prijs van zonlicht te voorzien; maar hij klaagde dat zijn voorraad zo gering was en hij verzocht mij hem iets te geven als aanmoediging van het vernuft, vooral omdat het een heel duur komkommerjaar was geweest; ik gaf hem een kleinigheid, want mijn gastheer had mij met het oog hierop van geld voorzien, daar hij hun gewoonte kende om ieder die hen kwam bezoeken iets af te bedelen.

* * *

Ik had tot dusver slechts de ene kant van de academie gezien. De andere is bestemd voor de beoefenaars der bespiegelende wetenschappen, en ook daarover zal ik het een en ander vertellen, maar eerst moet ik het nog hebben over een vermaarde persoonlijkheid, die men daar de „universele geleerde" noemt. Hij vertelde ons, dat hij al dertig jaar zijn gedachten had gewijd aan de veraangenaming van het menselijk bestaan.

* * *

De hele verhandeling was met grote scherpzinnigheid geschreven, en behelsde vele opmerkingen die voor staatslieden zowel merkwaardig als nuttig zijn. Maar naar mijn idee was zij niet helemaal volledig. Ik was zo vrij dit de schrijver mee te delen en bood aan hem desgewenst nog enige aanvulling te leveren. Hij ging hierop met meer bereidwilligheid in dan bij schrijvers gebruikelijk is, met name bij die van de plannenmakende soort, en verklaarde dat hij gaarne nadere inlichtingen zou krijgen.

* * *

De professor uitte zijn grote erkentelijkheid voor deze opmerkingen en beloofde dat hij er in zijn verhandeling met waardering van gewagen zou.

* * *

Ik zag in dat land niets dat mij tot langer verblijf kon noden, en begon over terugkeer naar Engeland te denken.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Het complex Waalre van Philips' Natuurkundig Laboratorium te Eindhoven.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Theoretische natuurkunde en industrieel onderzoek

H. B. G. Casimir

Inleiding

„Als je Casimir ontmoet, doe hem dan mijn groeten en spreek hem aan met Herr Direktor; daar kan hij niet tegen, dat weet ik". Met deze of soortgelijke woorden placht mijn grote leermeester Wolfgang Pauli — gedurende het academisch jaar 1932-33 ben ik zijn assistent geweest in Zürich — zo nu en dan fysici te instrueren die Nederland gingen bezoeken. En tijdens één van zijn laatste bezoeken aan ons land heeft hij zijn gevoelens zelfs nog duidelijker vertolkt. We zaten toen met een kleine vriendenkring bij elkaar, waarop iemand me vroeg of ik het niet erg moeilijk had gehad in Zürich. Daarop heb ik toen zo ongeveer als volgt geantwoord: „Neen, echt moeilijk niet. Pauli was over het algemeen niet zo streng. Hij was er juist met grote moeite in geslaagd zijn rijbewijs te halen, en omdat hij toen ongetrouwd was reed hij dan 's avonds nog wel eens met me de stad uit om in een landelijk restaurant rustig wat te gaan eten. We hadden de stilzwijgende afspraak dat zolang ik me maar zou onthouden van kritisch commentaar op zijn rijkunst, hij mijn fysica niet zou aanvallen. Nu wil ik niet over mijn eigen fysica opscheppen, maar die was toch bepaald beter dan Pauli's rijden, en daardoor stond ik sterk". Even zag je Pauli nadenken over mijn verhaal, en toen zei hij: „Ja, zo zal het wel geweest zijn, maar tegenwoordig rijd ik niet meer en jij doet geen fysica meer. Die Sache stimmt noch immer". Toen de hoofdredacteur van dit tijdschrift mij voorstelde iets te schrijven over de ervaringen van een theoretisch fysicus in een industrieel laboratorium schoot deze anekdote me weer te binnen.

Ik veronderstel dat Pauli's werkelijke gevoelens over mijn Herr-Direktor-schap van een industrieel researchlaboratorium een mengeling van verbazing, geamuseerdheid en spijt zijn geweest. Verbazing omdat iemand met enig begrip voor theorie en wat wiskundige capaciteiten, die in de oorlogsjaren zijn toevlucht had gezocht in een industrieel laboratorium, toen de oorlog voorbij was daar verkoos te blijven, in plaats van naar de universiteit terug te keren. Geamuseerdheid vanwege het feit dat zijn vroegere assistent zich blijkbaar wist te handhaven in zo'n weinig passende situatie. Spijt omdat ik wellicht werk zou hebben kunnen doen dat volgens zijn maatstaven veel waardevoller was. We moeten toegeven: een industrieel laboratorium is niet de plaats voor werk dat als Fundamenteel met een hoofdletter F mag worden gekenschetst, zoals ik al eens eerder heb gezegd. De relativiteitstheorie, de

Prof. Dr. W. Pauli (1900-1958)

kosmologie, de quantummechanica, de quantumelektrodynamica, de precieze theorie van fase-overgangen inclusief de orde-wanorde-overgangen, de veldentheorie en de theorie der elementaire deeltjes — geen van deze dingen is uit een industrieel laboratorium voortgekomen. Evenmin is de industrie erg snel in het verwerken van dit soort theoretische beschouwingen; vele van de diepgaandste en fraaiste resultaten hebben tot dusver geen enkele toepassing gevonden.

Toch heb ik nimmer het gevoel gehad dat een theoretisch fysicus in een industrieel laboratorium van zichzelf vervreemdt. Hieraan zitten drie kanten. Ten eerste verschaft de fundamentele theorie een eenvoudig en bevredigend systeem met behulp waarvan men de rijke overvloed aan experimentele en theoretische problemen en prestaties der moderne technologie kan blijven overzien. Dit is voor iemand die met een grote verscheidenheid van problemen heeft te maken, van bijzonder belang: zonder de steun die theoretisch inzicht geeft, zou hij maar al te spoedig verstrikt raken in velerlei details. Mijns inziens is voor iedere researchleider een degelijke kennis van de fundamentele theorie

Prof. dr. H. B. G. Casimir, tot 15 juni 1972 lid van de Raad van Bestuur van de N.V. Philips' Gloeilampenfabrieken en algemeen leider van de research van het Philips-concern.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

een eerste vereiste. Anders gezegd: de theorie is voor hem een onmisbaar stuk gereedschap.

Ten tweede is het misschien wel waar dat de problemen die de industrie oplevert niet zeer diepzinnig aandoen, maar ze zijn soms toch heel moeilijk en uitdagend. Aan die uitdaging zit nog een speciaal kantje. In het zogenaamde zuiver wetenschappelijke onderzoek hebben we nog wel eens de neiging alleen die problemen aan te pakken waarvan we geloven ze te kunnen oplossen, en we vermijden dan die welke er niet veelbelovend uitzien. In de industrie heeft men veel minder de gelegenheid om aan deze lafhartige neiging toe te geven.

Ten derde komt het voor dat ogenschijnlijk laag-bij-de-grondse zaken terugvoeren naar onverwacht fundamentele gedachten — zij het misschien geen Fundamentele.

In het volgende wil ik proberen de beschreven situatie te verduidelijken aan de hand van een aantal voorbeelden. Zij hebben betrekking op de theorie van het elektromagnetisme en op toepassingen van denkbeelden uit de groepentheorie.

Elektromagnetisme

De kennis van de theorie van Maxwell en van de verdere ontwikkeling daarvan door Lorentz maakt iemand nog niet tot elektrotechnisch ingenieur, maar het stelt hem wel in staat de essentiële eigenschappen van een willekeurig elektrisch of elektronisch apparaat of schakelingselement te begrijpen. Ik zou hier een versie van de theorie willen presenteren waarmee ik altijd zeer goed ben uitgekomen en waaraan ik boven elke andere formulering de voorkeur geef. Als eenhedensysteem gebruik ik dat van Gauss. In vacuüm geldt dan:

$$\operatorname{div} \boldsymbol{E} = 4\pi\varrho, \qquad \operatorname{div} \boldsymbol{B} = 0,$$

$$\operatorname{rot} \boldsymbol{B} = \frac{4\pi}{c}\boldsymbol{i} + \frac{1}{c}\dot{\boldsymbol{E}}, \qquad \operatorname{rot} \boldsymbol{E} = -\frac{1}{c}\dot{\boldsymbol{B}}.$$

Dezelfde vergelijkingen zijn van toepassing in materie, maar de ladingsdichtheid bevat dan ook een bijdrage afkomstig van de polarisatie, en in de stroomdichtheid doen atomaire stromen mee die corresponderen met veranderende polarisatie en met magnetisme. We mogen schrijven:

$$\varrho_{\text{totaal}} = \varrho_{\text{uitw}} + \varrho_{\text{pol}},$$

$$\boldsymbol{i}_{\text{totaal}} = \boldsymbol{i}_{\text{geleiding}} + \boldsymbol{i}_{\text{pol}} + \boldsymbol{i}_{\text{magn}}.$$

Hierdoor ontstaan de vergelijkingen:

$$\operatorname{div} \boldsymbol{E} = 4\pi(\varrho_{\text{uitw}} + \varrho_{\text{pol}}),$$

$$\operatorname{rot} \boldsymbol{B} = \frac{4\pi}{c}(\boldsymbol{i}_{\text{geleiding}} + \boldsymbol{i}_{\text{pol}} + \boldsymbol{i}_{\text{magn}}) + \frac{1}{c}\dot{\boldsymbol{E}}.$$

De ontbinding van de totale ladingsdichtheid en de totale stroomdichtheid in „uitwendige" en „inwendige" (atomaire) componenten hoeft niet altijd ondubbelzinnig te zijn, al is er dikwijls maar één keus die voor de hand ligt. We kunnen schrijven:

$$\varrho_{\text{pol}} = -\operatorname{div} \boldsymbol{P},$$

en nader onderzoek heeft geleerd dat deze vector $\boldsymbol{P}$ dikwijls in zeer goede benadering de gemiddelde waarde van de atomaire dipoolmomenten voorstelt. Dan geldt ook:

$$\boldsymbol{i}_{\text{pol}} = \dot{\boldsymbol{P}} \qquad \text{en} \qquad \boldsymbol{i}_{\text{magn}} = c \operatorname{rot} \boldsymbol{M}.$$

Soms, speciaal wanneer de keuze van ϱ_{pol} en $\boldsymbol{i}_{\text{magn}}$ ondubbelzinnig is, heeft het zin om twee hulpgrootheden in te voeren:

$$\boldsymbol{D} = \boldsymbol{E} + 4\pi\boldsymbol{P}, \qquad \boldsymbol{H} = \boldsymbol{B} - 4\pi\boldsymbol{M}.$$

Dan geldt namelijk:

$$\operatorname{div} \boldsymbol{D} = 4\pi\varrho_{\text{uitw}}, \qquad \operatorname{rot} \boldsymbol{H} = \frac{4\pi}{c}\boldsymbol{i}_{\text{geleiding}}.$$

Vanzelfsprekend kan men de theorie van het elektromagnetisme ook uitdrukken in de eenheden van het Giorgi-systeem, maar dan heeft men het bezwaar dat de vectoren $\boldsymbol{H}$ en $\boldsymbol{D}$ reeds optreden in het deel van de theorie dat voor het vacuüm geldt. Dit maakt het noodzakelijk nog sterker de nadruk te leggen op het feit dat deze vectoren hulpgrootheden zijn waarvan de fysische betekenis dikwijls niet ondubbelzinnig vastligt.

Enige voorbeelden daarvan:

a) Binnen een supergeleidende bol geldt $\boldsymbol{B} = 0$. Hoe zit het met $\boldsymbol{H}$? We kunnen de afschermende stroom opvatten als een zuivere geleidingsstroom. Dan is $\boldsymbol{H}$ ook binnen de bol gelijk aan nul. Als alternatieve mogelijkheid kunnen we stellen $\boldsymbol{i} = c \operatorname{rot} \boldsymbol{M}$. Dan vinden we voor een bol in een homogeen veld $\boldsymbol{B}_0$ dat $\boldsymbol{H} = \frac{3}{2}\boldsymbol{B}_0$ is.

b) In een diëlektricum dat bestaat uit een isolerende kunststof, zijn een groot aantal kleine geleidende bolletjes met straal r gevat. We kunnen ladingen op de bolletjes opvatten als uitwendige ladingen. Maar we kunnen ook schrijven:

$$\boldsymbol{P} = nr^3\boldsymbol{E},$$

waarin n het aantal bolletjes per cm³ voorstelt.

c) Een golf met hoekfrequentie ω loopt door een plasma dat n elektronen per cm³ bevat. Voor de elektronen geldt:

$$m\ddot{x} = eE,$$

wat leidt tot een stroom

$$\boldsymbol{i} = ne\dot{x} = -\frac{ne^2}{m\omega^2}\dot{\boldsymbol{E}}.$$

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Nu kunnen we hetzij schrijven:

$$\boldsymbol{i} = \dot{\boldsymbol{P}}, \qquad \text{met} \qquad \boldsymbol{P} = -\frac{ne^2}{m\omega^2}\boldsymbol{E},$$

hetzij, als we nog gebruik maken van $\dot{\boldsymbol{E}} = c \operatorname{rot} \boldsymbol{B}$:

$$\boldsymbol{i} = c \operatorname{rot} \boldsymbol{M}, \qquad \text{waarin} \qquad \boldsymbol{M} = -\frac{ne^2}{m\omega^2}\boldsymbol{B}.$$

De moraal van deze voorbeelden: bij twijfel doet men er goed aan de atomaire ladingen en stromen te beschouwen. Dit geldt vooral wanneer men zich bezighoudt met randvoorwaarden [1].

Ondanks de voorgaande opmerkingen, hebben $\boldsymbol{D}$ en $\boldsymbol{H}$ als hulpgrootheden toch dikwijls veel nut. Neem het geval van een permanente magneet. Geleidingsstromen zijn er niet, zodat rot $\boldsymbol{H} = 0$ is. Bijgevolg geldt (een elementaire stelling uit de vectoranalyse):

$$\int \boldsymbol{H}\cdot\boldsymbol{B}\,\mathrm{d}V = 0,$$

of

$$\left|\int \boldsymbol{H}\cdot\boldsymbol{B}\,\mathrm{d}V\right|_{\text{binnen magneet}} = \left(\int \boldsymbol{H}\cdot\boldsymbol{B}\,\mathrm{d}V\right)_{\text{buiten magneet}}.$$

Probeer nu te bereiken dat $\boldsymbol{H}\cdot\boldsymbol{B}$ binnen de magneet overal de maximale waarde aanneemt. Dan geldt:

$$\left(\int \boldsymbol{H}\cdot\boldsymbol{B}\,\mathrm{d}V\right)_{\text{buiten magneet}} = (\boldsymbol{B}\cdot\boldsymbol{H})_{\max} V_{\text{magneet}}.$$

Verder mag men stellen:

$$\left(\int \boldsymbol{H}\cdot\boldsymbol{B}\,\mathrm{d}V\right)_{\text{buiten magneet}} = \left(\int \boldsymbol{B}\cdot\boldsymbol{B}\,\mathrm{d}V\right)_{\text{luchtspeet}} + \\ + \left(\int \boldsymbol{B}\cdot\boldsymbol{B}\,\mathrm{d}V\right)_{\text{spreidingsveld}} + \left(\int \boldsymbol{H}\cdot\boldsymbol{B}\,\mathrm{d}V\right)_{\text{weekijzer}}.$$

Natuurlijk vertelt deze eenvoudige wiskundige beschrijving ons niet hoe we spreidingsvelden moeten berekenen. Zij vertelt ons evenmin hoe we een magnetisch circuit moeten ontwerpen om te bereiken dat $\boldsymbol{B}\cdot\boldsymbol{H}$ inderdaad zijn maximale waarde heeft. En ook zegt de beschrijving totaal niets over de manier waarop $(\boldsymbol{B}\cdot\boldsymbol{H})_{\max}$ nu afhangt van de structuur van het materiaal. Maar wel geeft zij ons een simpele en directe classificatie van deze problemen.

Bij toepassingen van de vergelijkingen van Maxwell in de praktijk is het vaak het gemakkelijkst deze te gebruiken in de integraalvorm. Dit is algemeen bekend: ik hoef alleen maar transformators en elektromagneten te noemen. Andere voorbeelden zijn nog het elektrische veld in een betatron, in de versnelsecties van een cosmotron, of langs de bundel in een klystron: men is gemakkelijk in staat de magnetische flux te bepalen die door een geschikt gekozen integratieweg wordt omvat. Velen, echter, hebben er wat moeite mee de differentiaalvergelijkingen om te zetten in iets tastbaars. Men beschouwe een elektromagneet met een symmetrievlak waarin $H_x = H_y = 0$ en H_z variabel is. We hebben dan:

$$\frac{\partial H_z}{\partial x} = \frac{\partial H_x}{\partial z}.$$

Indien H_z afneemt in de richting van positieve x, zullen we een negatieve H_x vinden wanneer we ons buiten het symmetrievlak in de richting van H_z begeven. De krachtlijnen zijn zo gebogen dat het krommingsmiddelpunt ligt in de richting van toenemende H_z. Dit is essentieel in de theorie voor de baanstabiliteit in het verticale vlak bij cyclotrons en betatrons.

Een veel gebruikt theorema zegt dat wanneer men een grootheid heeft die voldoet aan de vergelijking van Laplace, men door middel van deze grootheid over het oppervlak van een bol de waarde ervan in het middelpunt van de bol vindt. (Om dit te bewijzen ontwikkele men naar bolfuncties; alle bolfuncties van de orde $\geqslant 1$ leveren een bijdrage 0; slechts de constante term blijft over.) Deze stelling kan bij sommige numerieke methodes gebruikt worden. Hij speelt ook een rol bij het in kaart brengen van een magneetveld: een homogeen gewikkeld bolvormig proefspoeltje meet exact de waarde van het veld in zijn middelpunt, hoe inhomogeen het veld ook moge zijn.

Terloops vermeld ik hier nog een grappige fysische interpretatie van het feit dat het gemiddelde van de bolfunctie van de tweede orde nul is: $\frac{1}{2}\overline{(3\cos^2\varphi - 1)} = 0$. Wanneer twee dipolen die zich in willekeurige posities bevinden, door een homogeen magneetveld gericht worden, dan zal hun gemiddelde wisselwerkingsenergie nul zijn indien dit veld al roterend alle richtingen doorloopt. Ik heb voorgesteld dit verschijnsel te gebruiken voor het dispergeren van magnetische poeders, maar ik geloof toch niet dat het een erg praktisch voorstel is.

Op deze wijze zou nog menige bladzijde gevuld kunnen worden — misschien ga ik dat in het kader van een boek nog eens doen —, maar dat is niet mijn doel bij deze gelegenheid. Ik heb alleen willen laten zien dat men bij elke stap in een industrieel laboratorium, zoals Philips' Natuurkundig Laboratorium, de theorie van het elektromagnetisme op frappante wijze geïllustreerd vindt.

En men ontmoet niet alleen maar eenvoudige gevallen. Er zijn vele problemen die de wiskundige gaven van een fysicus en soms ook de capaciteit van een grote computer en de kundigheid van zijn programmeur op een zware proef stellen. In de loop der jaren hebben de Philips-laboratoria veel tot de oplossing van dit soort problemen bijgedragen. Ik mag hier misschien het werk

[1] Men zie H. B. G. Casimir, in: Polarisation, Matière et Rayonnement, Volume jubilaire en l'honneur d'Alfred Kastler, Presses Universitaires de France 1969, p. 185, en H. B. G. Casimir, Philips Res. Repts. 21, 417, 1966.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

van Bremmer en Van der Pol vermelden over de buiging van radiogolven die om de aarde lopen, of het werk van Bouwkamp over antennes en over buigingsverschijnselen bij gaten. Ik zou nog eens lof willen toezwaaien aan ir. G. de Vries, die steeds de resultaten van vele vernuftige berekeningen op gulle wijze aan zijn collega's ter beschikking stelde zonder dat hij er zich om bekommerde zijn resultaten ook te publiceren.

Hiermee samenhangend zou ik nog eens kort de theorie van wervelstromen in ferromagnetische materialen willen noemen. Wijlen dr. Snoek had erop gewezen dat er een grensgeval bestaat waarbij de dikte van de skinlaag klein is ten opzichte van de dimensies d van het materiaalmonster, terwijl het produkt μd een willekeurig grote waarde kan hebben. Daarop vond ik toen [2] dat in dat geval het veld afgeleid kan worden uit een oplossing van de vergelijking van Laplace met de merkwaardige randvoorwaarde

$$\frac{\partial \Phi}{\partial n} - \frac{\mu d}{1+\mathrm{j}} \left[\frac{\partial^2 \Phi}{\partial n^2} + \left(\frac{1}{R_1} + \frac{1}{R_2} \right) \frac{\partial \Phi}{\partial n} \right] = 0,$$

waarin R_1 en R_2 de hoofdkromtestralen voorstellen.

Wat de derde kant aan theoretisch werk in een industrieel research-laboratorium aangaat, hoop ik dat men mij niet euvel duidt dat ik uit persoonlijke ervaringen put. Bij het bestuderen van staande golven in trilholten voor cm-golven [3] heb ik geleerd dergelijke trillingswijzen als zeer reële fysische zaken te beschouwen, veel reëler dan de formele stralingstheorie ze me ooit had doen zien. Dit onderzoek heeft me ook eenvoudige formules voor de behandeling van kleine storingen verschaft. En dat gaf me de moed de nulpuntsenergie ($\Sigma \frac{1}{2} h\nu$) van deze trillingswijzen te bestuderen. Deze som divergeert sterk, maar het is mogelijk een ondubbelzinnig gedefinieerde waarde toe te kennen aan het verschil tussen de (ongedefinieerde) waarden van deze som die onder verschillende condities optreden. Dit heeft geleid tot de voorspelling [4] van een universele wisselwerking U van aantrekkend karakter tussen twee metalen platen op afstand a, ieder met een oppervlakte l^2:

$$U = -\frac{\pi^2}{720} \hbar c \frac{l^2}{a^3}.$$

Experimenten hebben deze voorspelling bevestigd. Het werk heeft ook geleid tot een vereenvoudigde afleiding [5] van de formules voor de geretardeerde Van-der-Waals-krachten, gevonden door Polder en mij [6] in antwoord op een vraag van Overbeek, die was voortgekomen uit het werk van laatstgenoemde aan de stabiliteit van colloïdale suspensies.

Dit ene voorbeeld zou met vele andere aangevuld kunnen worden.

Groepentheorie

Mijn tweede reeks voorbeelden draait om denkbeelden uit de groepentheorie, d.i. de theorie van de structuur en de representatie van bepaalde verzamelingen van transformaties. De formele theorie heeft een belangrijke rol gespeeld bij de ontwikkeling van de quantummechanica; de algemene ideeën erachter verschaffen een voortreffelijk patroon van eenheid scheppend inzicht. Deze tak van de wiskunde zal helaas vele lezers wellicht niet zo vertrouwd zijn, en hier is het eigenlijk niet de juiste plaats voor verklarende details. Ik hoop maar dat mijn wat oppervlakkige aanduidingen de strekking voldoende bloot leggen.

Eenheden en commutatieve groepen

In de fysica hebben we te maken met twee soorten vergelijkingen. Aan de ene kant zijn er de wiskundige identiteiten, zoals

$$\sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} = x \prod_{n=1}^{\infty} \left(1 - \frac{x^2}{n^2 \pi^2} \right),$$

die gelden voor willekeurige (reële of complexe) waarden van x. Aan de andere kant bestaan er vergelijkingen die de eigenschappen van de natuur beschrijven. Neem als voorbeeld de vergelijking

$$c^2 = a^2 + b^2, \tag{1}$$

waarvan men kan zeggen dat hij een eigenschap van een rechthoekige driehoek in een plat vlak beschrijft, een interessant feitelijk gegeven over het kwadraat van de schuine zijde. In deze vergelijking stellen de letters a, b en c lengten voor die gemeten zijn met behulp van een zekere eenheid van lengte. Indien we deze eenheid met een factor $1/\alpha$ vermenigvuldigen, zullen de getallen a, b en c een transformatie ondergaan:

$$a' = \alpha a, \quad b' = \alpha b, \quad c' = \alpha c.$$

Al de mogelijke transformaties van deze soort vormen nu samen een commutatieve groep (of Abel-groep), en men noemt verg. (1) invariant voor deze groep.

Wanneer we zeggen dat het volume V van een bol gegeven wordt door

$$V = (4\pi/3)\, R^3, \tag{2}$$

dan is dit waar wanneer we het volume meten met behulp van de eenheid van lengte tot de derde macht. Verg. (2) is invariant voor een transformatie

$$V' = \alpha^3 V, \quad R' = \alpha R.$$

Maar we zouden ook een onafhankelijke eenheid van volume zoals de (Engelse) gallon of de (Amerikaanse) barrel of desnoods de paplepel kunnen invoeren. Dan moeten we schrijven:

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

$$V = v_0 \frac{4\pi}{3} R^3,$$

waarin v_0 het volume is van de kubus met ribben gelijk aan eenheid van lengte. Bij deze vergelijking is een grotere groep van transformaties toegestaan:

$$V' = \beta V, \qquad R' = \alpha R, \qquad v_0' = (\beta/\alpha^3)v_0.$$

Dit voorbeeld bewijst dat we een zekere mate van vrijheid genieten: we kunnen onze grootheden op verschillende manieren definiëren en afhankelijk van deze definities zullen onze vergelijkingen invariant worden voor een meer of minder beperkte groep van transformaties. In het algemeen mag men de keuze maken die het best uitkomt. Een interessant voorbeeld is het geval van de elektromagnetische eenheden [7].

Abel-groepen; lineaire verschuiving

Een kristalrooster met roostervectoren $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, $\boldsymbol{u}_3$ laat een verschuivingsoperatie toe van de algemene gedaante

$$D = \boldsymbol{u}_1^{n_1} \boldsymbol{u}_2^{n_2} \boldsymbol{u}_3^{n_3},$$

waarbij de vectoren $\boldsymbol{u}_i$ gecommuteerd mogen worden. De verzameling D's vormt een oneindige (maar wel aftelbare) Abel-groep. Hieruit volgt (theorema van Bloch) dat de golffunctie van een elektron dat zich binnen een dergelijk rooster beweegt, de volgende vorm heeft:

$$\psi = e^{jkx} p(x), \qquad (3)$$

waarbij $p(x)$ periodiek is: $p(Dx) = p(x)$.

Een magnetron met N trilholten laat rotaties R, R^2, . . . toe, waarbij $R^N = E$. Hieruit volgt dat het veld een superpositie van trillingswijzen is waarbij het veld in trilholte m gegeven wordt door

$$F_m = e^{2\pi jqm/N} f_q(x), \text{ met } q = 0,1, \ldots, N-1. \qquad (4)$$

Hierin hangt $f_q(x)$ wel van q af, maar is in alle trilholten gelijk.

De overeenkomst tussen de beide voorbeelden is wel zeer duidelijk: in beide gevallen is de situatie invariant voor een bepaalde commutatieve groep; in het eerste geval zijn in de elementen van de groep de verschuivingen beschreven met de roostervector $n_1\boldsymbol{u}_1 + n_2\boldsymbol{u}_2 + n_3\boldsymbol{u}_3$, in het tweede zijn de elementen rotaties over een hoek $m.\ 2\pi/N$; dit bepaalt de algemene vorm van het veld.

Soortgelijke beschouwingen gelden voor allerlei lineaire ketens, voor elastische trillingen in een kristal of in een benzeenmolecuul, enz., enz.

De groep der rotaties

Het vectormodel voor de atomen kan men op een elegante wiskundige wijze verklaren door de theorie der irreducibele representaties van de driedimensionele rotatiegroep toe te passen op de atomaire golffuncties. Het is nuttig om te weten dat in de theorie van het elektromagnetisme de bolfuncties een basis vormen voor multipoolstralingsvelden [8], maar er is dan toch wel een verschil met het scalaire geval. In het geval van een scalaire potentiaal is een $(2l + 1)$-dimensionale representatie noodzakelijkerwijs voorzien van de pariteitsfactor $(-1)^l$. Bij een vectorveld, het andere geval, kan de pariteitsfactor voor een $(2l + 1)$-dimensionale representatie zowel $+1$ als -1 zijn. In het stralingsveld van een elektrische 2^l-pool draagt de elektrische golf $\boldsymbol{E}$ de pariteitsfactor $(-1)^l$, de magnetische $\boldsymbol{H}$ de factor $(-1)^{l+1}$; bij een magnetische 2^l-pool is het net andersom.

Dit soort groepentheoretische beschouwingen kan de behoefte aan gedetailleerde berekeningen niet opheffen. Men kan echter wél met de groepentheorie vele algemene eigenschappen van de expliciete oplossingen der berekeningen voorspellen. Dit leidt tot belangrijke vereenvoudigingen en het verschaft een nuttige controle op die resultaten.

Symmetrie en integralen van de bewegingsvergelijkingen

Wanneer een bepaalde hamiltoniaan een continue groep transformaties toelaat, dan zijn de grootheden waarmee de infinitesimale canonieke transformaties van deze groep gegenereerd kunnen worden, integralen van de bewegingsvergelijkingen. Dit nuttige en diepgaande theorema geldt zowel in de quantummechanica als in de klassieke mechanica; ons eerste voorbeeld behoort beslist tot de klassieke. Men aanschouwe een schoonrijdster op het ijs. Het krachtveld waarin zij zich bevindt is symmetrisch om de verticale as. Ook het ijs voldoet aan deze symmetrie. Bijgevolg is de grootheid voor het genereren van de infinitesimale transformaties die corresponderen met rotaties om de verticale as, een integraal van de beweging: het impulsmoment. Omdat nu het impulsmoment het produkt is van hoeksnelheid en traagheidsmoment, kan de hoeksnelheid beheerst worden door het traagheidsmoment te veranderen. Een schoonrijdster die pirouetteert met een steeds groter wordende snelheid, geeft niet alleen een vertoning van élégance maar ook een proefondervindelijk bewijs van het isotrope karakter van de ruimte. Soortgelijke beschouwingen zijn van toepassing op het naar Einstein en De Haas genoemde

[2] H. B. G. Casimir, Philips Res. Repts. **2**, 42, 1947.
[3] H. B. G. Casimir, Philips Res. Repts. **6**, 162, 1951.
[4] H. B. G. Casimir, Proc. Kon. Ned. Akad. Wetensch. **51**, 793, 1948.
[5] H. B. G. Casimir, J. Chimie phys. **46**, 407, 1949.
[6] H. B. G. Casimir en D. Polder, Phys. Rev. **73**, 360, 1948.
[7] H. B. G. Casimir, On electromagnetic units, Helv. Phys. Acta **41**, 741-742, 1968. Zie ook dit nummer blz. 359.
[8] H. B. G. Casimir, Helv. Phys. Acta **33**, 849, 1960.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

effect [*]; ook daarbij blijft het totale impulsmoment behouden.

Er bestaat een analoge betrekking tussen lineaire verschuivingen en behoud van impuls; hierover behoeft niet veel te worden gezegd. Maar een veel interessantere situatie ontstaat wanneer we verschijnselen bestuderen in een kristal dat zich in rust bevindt. Zijn de hamiltonianen die we gebruiken invariant voor een infinitesimale translatie ten opzichte van het kristal? Als ze dat inderdaad zijn, moet er een corresponderende „pseudo-impuls" bestaan, die behouden blijft; zijn ze niet invariant, dan mag zulk een conclusie niet getrokken worden. En het antwoord luidt alleen maar bevestigend wanneer het kristal als een continuüm kan worden opgevat. Beschouw nu eens elastische golven. In een continuüm is er een constante pseudo-impuls, wat ook het niet-lineaire verschijnsel moge zijn dat tot wisselwerking tussen de golven leidt. Een kristal is echter *geen* ideaal continuüm; bijgevolg blijft de pseudo-impuls niet exact behouden. Dit is de kern van de theorie van Peierls over warmtegeleiding; zijn „Umklapp-Prozesse" brengen het verschil tussen een ideaal continuüm en een kristalrooster tot uitdrukking.

Omkering van de tijd

De vergelijkingen van de elektrodynamica en de mechanica zijn invariant voor de substitutie $t \rightarrow -t$. Dit is zowel waar voor de klassieke theorie als voor de quantumtheorie. Maar macroscopisch is het gedrag van de materie bepaald niet reversibel, ofschoon de materie toch uit atomen bestaat waarvoor invariantie voor tijdsomkering wèl moet gelden. Het antwoord op deze schijnbare tegenspraak wordt gegeven door de statistische mechanica. Onsager heeft aangetoond dat, ofschoon de paradox kan worden opgeheven, de omkeerbaarheid van de atomaire processen tot uiting komt in bepaalde macroscopische symmetrie-eigenschappen.

Onsagers theorie heeft een onverwacht verband gelegd tussen Tellegens beschouwingen over gyrators, de betrekkingen van Thomson tussen thermo-elektrische grootheden, de symmetrie van geleidingstensoren, en vele andere verschijnselen.

De bestudering van de betrekkingen tussen de netwerktheorie en de denkbeelden van Onsager heeft geleid tot een kleine uitbreiding van Onsagers formalisme [9], wat opnieuw een voorbeeld is van een geval waarin praktische toepassingen terugvoeren naar fundamentele zaken.

Slotopmerking

Dit artikel is misschien wel erg persoonlijk geworden. Bij mijn pogingen uiteen te zetten hoe een theoreticus reageert op de problemen van een industrieel laboratorium, heb ik wel wat veel van mijn eigen stokpaarden laten opdraven en teveel aandacht aan mijn eigen bijdragen geschonken, zodat de ideeën en de verrichtingen van anderen wat in het gedrang zijn gekomen.

Misschien kan men dit begrijpen. Maar voor ik eindig, wil ik nog eens onderstrepen dat in de industriële research het soort begrip dat verschaft wordt door de fundamentele theorie, slechts een eerste stap is, en bepaald niet de laatste. Slechts een lange weg van inspanning en volharding voert naar een nieuw produkt of een nieuw proces. Industriële innovatie vereist originaliteit, technologisch kunnen, nauwkeurigheid, uithoudingsvermogen en organisatorische capaciteiten, en voor dat alles wil ik mijn dank en bewondering uitspreken jegens mijn medewerkers in de research-laboratoria van het Philips-concern.

[*] Het Einstein-De-Haas-effect is het optreden van een klein koppel (impulsmoment) bij het magnetiseren (of demagnetiseren) van een ijzeren staaf. De aanleiding tot het effect vormt de elektronenspin, waarbij een magnetisch moment samengaat met een mechanisch impulsmoment. *(Red.)*

[9] H. B. G. Casimir, Rev. mod. Phys. 17, 343, 1945, of Philips Res. Repts. 1, 185, 1946.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Prof. Casimir bij het uitreiken van een prijs aan een groep deelneemsters aan de finale van de Europese Wedstrijd voor Jonge Onderzoekers, welke jaarlijks door Philips wordt georganiseerd in het Evoluon te Eindhoven.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

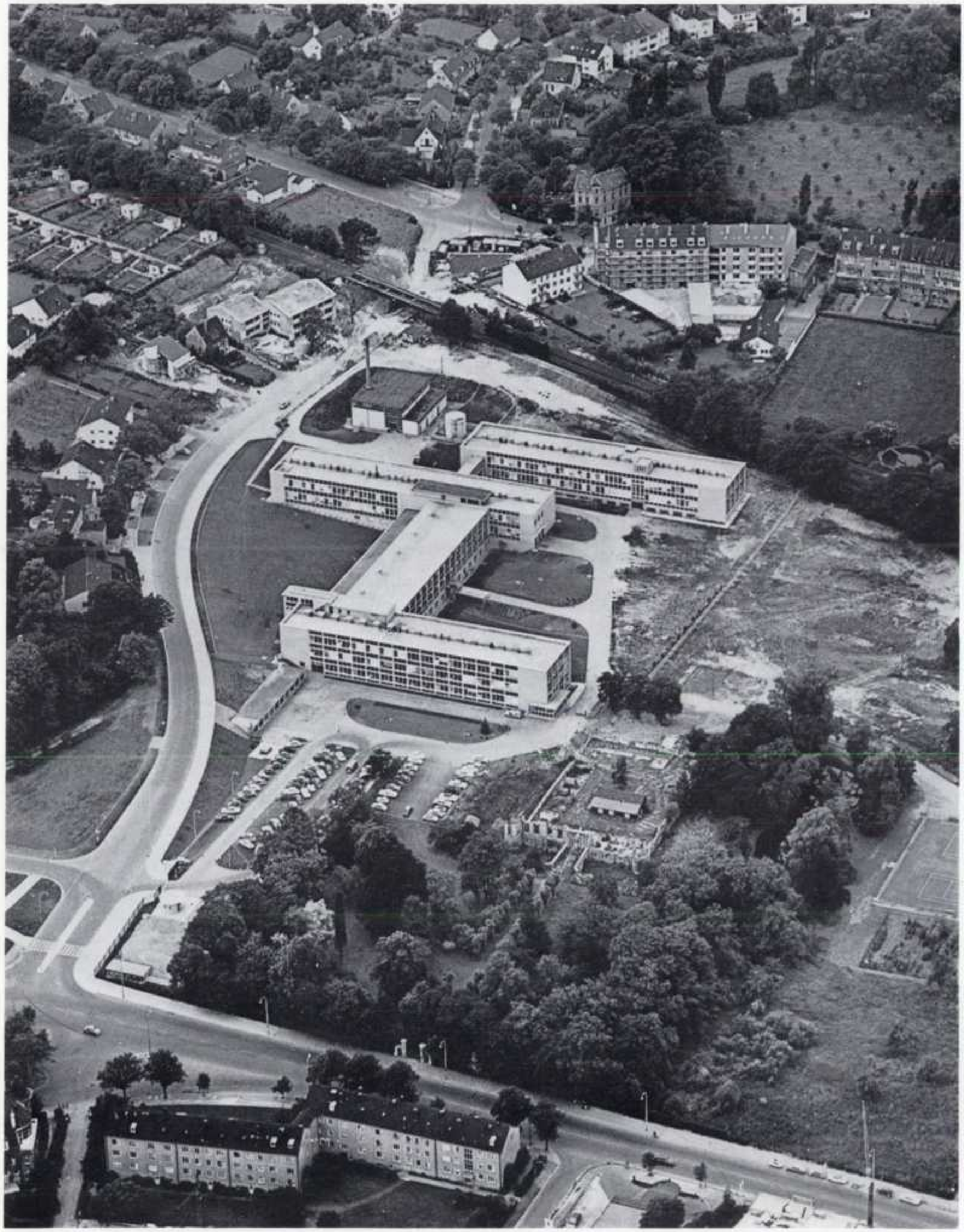


Foto Aero-Foto; Freigabe-Nr. Reg. Präs. Düsseldorf 831/06

Philips Forschungslaboratorium Aachen.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

De levensduur van gloeilampen

H. Hörster, E. Kauer en W. Lechner

Inleiding

Hoewel gloeilampen al een 60-tal jaren in massaproduktie worden vervaardigd, verkeert men nog goeddeels in het ongewisse over de fysische processen die zich hierin afspelen. Een vraag waarop men het antwoord nog schuldig blijft, is welk mechanisme de levensduur van de lamp beheerst. Het is wel algemeen bekend dat het doorbranden wordt veroorzaakt door het verdampen van wolfraam uit de gloeidraad, maar deze verklaring is ontoereikend. Dit blijkt uit een nadere beschouwing van een homogene draad die bij een constante spanning gloeit met een uniforme temperatuurverdeling, zodat de mate van verdamping langs de as overal gelijk is. Onder deze voorwaarden zou de diameter van de draad geleidelijk afnemen en de weerstand ervan naar evenredigheid toenemen, hetgeen een gestadige daling van de temperatuur tot gevolg zou hebben. Dit zou tot een oneindige levensduur leiden. Hiervan is evenwel geen sprake; de draad brandt immers op een of ander punt door na verloop van een eindige tijd, gedurende welke de verdampingsverliezen van de rest van de draad gering zijn.

Dit kan men verklaren door de aanwezigheid van kleine foutjes, de zogenaamde *hete plekken*, die een dusdanige invloed hebben op de energiebalans van de draad, dat de temperatuur ervan, en dus ook de verdamping, plaatselijk toeneemt. Dit heeft tot gevolg dat de diameter in sterkere mate afneemt; de stroom wordt immers voornamelijk beheerst door het foutloze deel van de draad. Op deze wijze wordt de plaatselijke stijging van de temperatuur van de hete plek versterkt als gevolg van de cyclus: verdamping → vermindering van de diameter → extra verdampingsverliezen, enz. R. Becker ging in 1925 van deze conceptie uit voor zijn kwalitatieve beschouwingen over de levensduur van een gloeidraad [1].

In het eerste deel van dit artikel worden, uitgaande van het model van de hete plek, kwantitatieve uitdrukkingen afgeleid voor de levensduur en voor andere belangrijke grootheden van een gloeidraad in vacuüm; hiertoe gaan we na hoe de fouten in de draad in het verloop van de tijd aangroeien [2].

In het tweede deel wordt het model toegepast op de met gas gevulde lamp. Hier wordt het transport van de massadeeltjes beheerst door de diffusie van wolfraamatomen door een edelgas. Voor een rechte draad behoudt het model zijn geldigheid, maar voor een gespiraliseerde draad treden sterke afwijkingen op, zoals een vermindering van de als „fataal gewichtsverlies" bekende grootheid. Dit verschijnsel maakt het nodig een nieuw model in te voeren, waarin rekening wordt gehouden met zowel radiale als axiale diffusie van het wolfraam [3]. We zullen deze beide grootheden vergelijken en dieper ingaan op het geval dat het axiale massatransport het radiale overtreft. Bij laagspanningslampen blijkt de levensduur door het axiale massatransport te worden bepaald [4].

Het doorbranden van een draad in vacuüm

Formulering van het model van de hete plek

Ter verklaring van het mechanisme dat voor het doorbranden van een gloeidraad verantwoordelijk is, veronderstellen we dat in de energiebalans van de draad een storing aanwezig is. Men kan dit proces in detail bestuderen door het gedrag van een lange, rechte draad te beschouwen. In het model (*fig. 1*) zij de temperatuur van de draad T_1, met uitzondering van een plaatselijke storing, waar zich een hete plek met een temperatuur T_2 bevindt. Deze temperatuur verschilt slechts

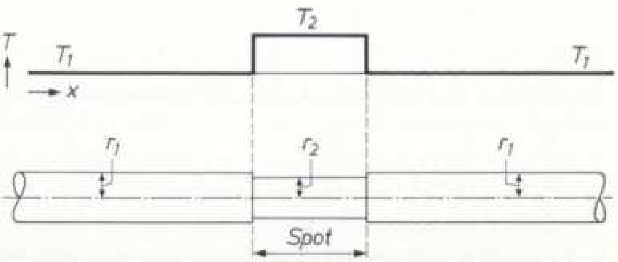


Fig. 1. Ter definitie van de plaatselijke onregelmatigheid in de draad en de „hete plek" (*Spot*). De temperatuur T is uitgezet als functie van de afstand x langs de draad. T_1 is de temperatuur van de draad, T_2 die van de hete plek. r_1 is de straal van de draad, r_2 die van de plek.

Prof. dr. E. Kauer, directeur van Philips Forschungslaboratorium Aachen GmbH (PFA), Aken, Duitsland; dr. H. Hörster en W. Lechner, medewerkers van PFA.

[1] R. Becker, Z. techn. Physik 6, 309, 1925.
[2] In het julinummer van J. of IES 1972 is over dit onderwerp een artikel van de auteurs verschenen. Tevens is op de jaarlijkse IES-conferentie, 17-20 augustus 1971, Chicago, Ill., U.S.A., een voordracht gehouden.
[3] Van dezelfde gedachtengang werd gewag gemaakt door F. Blau; hij wordt geciteerd door F. Koref en H. C. Plaut, Z. techn. Physik 11, 515, 1930.
[4] Zie ook B. A. Moys, Trans. Illum. Engng. Soc. 29, 13, 1964.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

weinig van T_1. In fig. 1 is aangenomen dat de fout wordt veroorzaakt door een insnoering van de draad.

Voor het beschrijven van de levensduur en andere belangrijke fysische grootheden moet men bepalen op welke wijze de draadtemperatuur T_1 en de temperatuur T_2 van de hete plek met de tijd veranderen. Hiertoe nemen we aan dat alle van belang zijnde grootheden uitsluitend aan de vermindering van de draaddiameter moeten worden toegeschreven. Deze vermindering wordt bepaald door verdampingsverliezen, die uitsluitend van de temperatuur afhankelijk zijn. Gaan we uit van de energiebalans van een foutloze draad, dan mogen we ter berekening van de temperatuur T_1 en van de stroom I als functie van de tijd t, aannemen dat de elektrische weerstand van de hete plek te verwaarlozen is ten opzichte van die van de gehele draad. Met andere woorden: de stroom I door een gloeidraad, aangesloten op een bron met constante spanning, is uitsluitend een functie van de elektrische weerstand van het foutloze deel van de draad, en dientengevolge van de veranderingen hiervan met de tijd. Dit betekent dat we het verloop van de temperatuur van de hete plek met de tijd kunnen berekenen indien $I(t)$, de stroom als functie van de tijd, bekend is. We zullen de levensduur τ definiëren als de tijd die is verlopen als de temperatuur $T_2(t)$ het smeltpunt van wolfraam heeft bereikt. Verder nemen we aan dat de warmtegeleiding tussen de hete plek en het foutloze deel van de draad geen invloed heeft op het verloop van de temperatuur van de plek, d.w.z. de lengte van de hete plek wordt verondersteld groot te zijn ten opzichte van de diameter ervan.

Het verloop van de temperaturen met de tijd kan worden berekend uit twee differentiaalvergelijkingen in T_1 en T_2. De oplossingen zijn vrij ingewikkeld en worden voornamelijk bepaald door de verdampingssnelheid $m(T)$, de draaddiameter r en het aanvankelijke verschil in temperatuur ΔT_0 tussen de hete plek en de foutloze draad. Ze hebben de volgende algemene vorm:

$$T_1(t) = \mathrm{f}\{m(T_{10}), r_{10}, T_{10}\}, \qquad (1)$$

$$T_2(t) = \mathrm{f}\{m(T_{10}), r_{10}, T_{10}, \Delta T_0\}, \qquad (2)$$

waarin r_{10} en T_{10} de straal en de temperatuur van de foutloze draad aan het begin van de levensduur voorstellen.

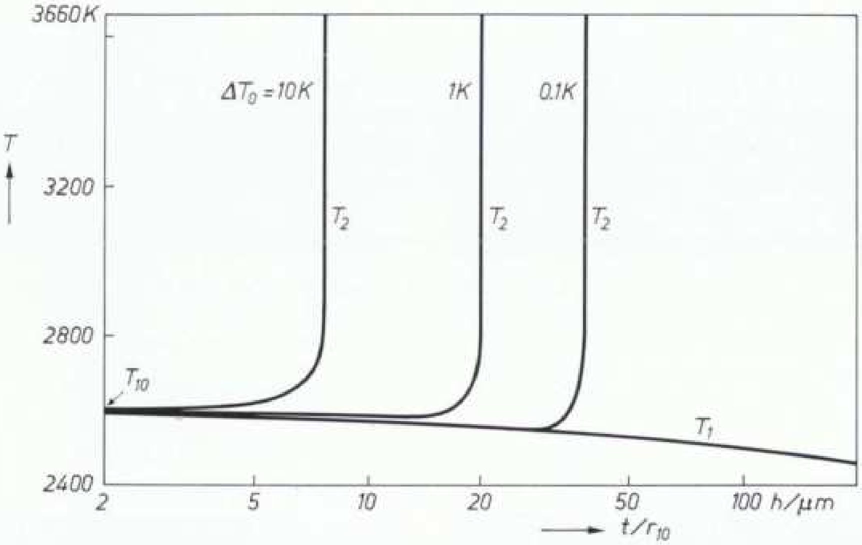


Fig. 2. Verloop als functie van de tijd van de temperatuur T_1 van een foutloze draad en van die van de hete plek, T_2, bij een constante spanning, voor het geval dat de beginwaarde T_{10} van T_1 2600 K bedraagt. Langs de abscis is uitgezet het quotiënt van de tijd t en de beginwaarde r_{10} van de straal van de draad. De drie krommen T_2 gelden voor drie waarden van het verschil ΔT_0 tussen de aanvankelijke temperatuur T_2 van de hete plek en de aanvankelijke temperatuur T_{10} van de draad.

In *fig. 2* zijn de oplossingen voor $T_1(t)$ en $T_2(t)$ voor een begintemperatuur van 2600 K logaritmisch uitgezet tegen het quotiënt t/r_{10}. De op deze wijze voorgestelde krommen gelden voor draden met een willekeurige straal. De parameter ΔT_0 stelt het verschil voor tussen de aanvankelijke temperatuur van de hete plek en die van de foutloze draad en is een maat voor de aanvankelijke fout. De vorm van $T_1(t)$ is in overeenstemming met hetgeen van een foutloze draad kan worden verwacht; de verdampingsverliezen veroorzaken een geleidelijke vermindering van de diameter, hetgeen bij een constante spanning een daling van de temperatuur tot gevolg heeft. Het verloop van de temperatuur T_2 als functie van de tijd wordt beheerst door het temperatuurverschil ΔT_0. Voor een gegeven waarde van ΔT_0, b.v. 0,1 K, blijkt de temperatuur $T_2(t)$ van de plek ge-

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

durende lange tijd nauwelijks van $T_1(t)$ af te wijken; daarna echter treedt er een sterke stijging op en wordt het smeltpunt van wolfraam (3655 K) snel bereikt. De draad heeft dan het einde van zijn levensduur τ bereikt. Fig. 2 toont duidelijk dat τ zeer sterk afhangt van het aanvankelijke temperatuurverschil ΔT_0.

Door in de formule voor de temperatuur $T_2(t)$ van de hete plek het smeltpunt van wolfraam in te vullen, verkrijgen we een analytische vergelijking voor de levensduur τ. Deze blijkt evenredig te zijn met de diameter van de draad en omgekeerd evenredig met de snelheid waarmee de foutloze draad verdampt; bovendien is de levensduur een functie van het aanvankelijke temperatuurverschil, ΔT_0. Het verband wordt gegeven door:

$$\tau = C\,\frac{r_{10}}{m(T_{10})}\left[\left\{1-\left(1+\frac{\Delta T_0}{T_{10}}\right)^{-a}\right\}^{-b}-1\right], \qquad (3)$$

waarin $m(T_{10}) = m_0 T_{10}{}^{\gamma}$ de verdampingssnelheid voorstelt, uitgedrukt als een machtsfunctie. Deze is een benadering van de exponentieel verlopende, door de temperatuur bepaalde verdampingsverliezen $m = m_0' \exp(-Q/RT)$, waarin Q, de verdampingswarmte, 205 kcal/mol bedraagt [5]. De benadering wordt verkregen door de afgeleiden van beide vergelijkingen (de machtsfunctie en de exponentiële functie) gelijk te stellen bij een bepaalde temperatuur. Zo verkrijgen we uit $\gamma = Q/RT$ en $T = T_{10} = 2600$ K, dat $\gamma = 40$ is, $a = \frac{1}{3}(4+3\gamma) = 41{,}3$ en $b = (\gamma-4)/(4+3\gamma) = 0{,}29$. C is een constante. Als het model juist is, dan betekent dit dus dat de levensduur uitsluitend afhangt van de begintemperatuur T_{10}, van de draaddiameter r_{10} en van de aanvankelijke relatieve fout $\Delta T_0/T_{10}$. Proeven hebben aangetoond dat de uitdrukking (3) inderdaad weergeeft op welke wijze de levensduur afhangt van de temperatuur en de draaddiameter.

Nu we een uitdrukking hebben gevonden voor de levensduur, kunnen we een analytische uitdrukking afleiden voor het „fatale gewichtsverlies" van een gloeidraad. Dit wordt gedefinieerd als de verhouding van de verdampingsverliezen ΔW van een gloeidraad gedurende zijn levensduur, tot zijn aanvankelijke massa W_0. Hiervoor geldt:

$$\frac{\Delta W}{W_0} = \frac{2\pi l}{W_0}\int_0^{\tau} r_1(t)\, m\{T_1(t)\}\,\mathrm{d}t, \qquad (4)$$

waarin l de lengte van de draad voorstelt.

In het verleden kon het fatale gewichtsverlies alleen proefondervindelijk worden vastgesteld. Het is echter een belangrijke grootheid, gezien zijn onafhankelijkheid van temperatuur en draaddiameter. Deze grootheid is voor een bepaald soort lamp (b.v. een vacuümlamp) constant binnen een begrensd gebied. Met behulp van (3) kunnen we uit (4) afleiden dat

$$\frac{\Delta W}{W_0} = 1-\left\{1-\left(1+\frac{\Delta T_0}{T_{10}}\right)^{-a}\right\}^{c}, \qquad (5)$$

waarin $c = 8/(4+3\gamma) \approx 0{,}065$ bij $\gamma = 40$. Het is verrassend dat $\Delta W/W_0$ uitsluitend wordt bepaald door de aanvankelijke relatieve fout $\Delta T_0/T_{10}$, en deze is in het van belang zijnde temperatuurgebied (2500 tot 3000 K) onafhankelijk van T_{10}. Vergelijking (5) is dus in overeenstemming met de waarneming dat draaddiameter en brandtemperatuur niet van invloed zijn, d.w.z. het fatale gewichtsverlies is onafhankelijk van het verloop van de stroom en spanning met de tijd. Voor het tot uitdrukking brengen van de kwaliteit is het dus een ideale grootheid, die de voorkeur verdient boven de levensduur. Deze hangt immers in sterke mate af van de brandtemperatuur T_{10}, en dus van de aangelegde spanning. Het gevolg daarvan is dat de invloed van ΔT_0 reeds door kleine veranderingen van de brandtemperatuur aan de waarneming wordt onttrokken.

Grootte van de aanvankelijke fouten

Men kan het fatale gewichtsverlies gemakkelijk meten. Proeven, uitgevoerd met 50 vacuümlampen (110 V, 75 W) gaven voor $\Delta W/W_0$ een gemiddelde waarde van 15%. Substitueert men deze waarde in (5), dan levert dit een temperatuurverschil ΔT_0 op van 5 K. De verrassend lage waarde van dit temperatuurverschil kan wellicht worden toegeschreven aan de aanwezigheid van een aantal onafhankelijke fouten die een plaatselijke verstoring van de energiebalans veroorzaken.

Uit de energiebalans volgt:

$$\frac{\Delta T_0}{T_{10}} = -\frac{3}{4}\left(\frac{\Delta r}{r}\right)_{\varrho,\varepsilon} + \frac{1}{4}\left(\frac{\Delta\varrho}{\varrho}\right)_{r,\varepsilon} - \frac{1}{4}\left(\frac{\Delta\varepsilon}{\varepsilon}\right)_{r,\varrho} + \ldots . \qquad (6)$$

Het relatieve temperatuurverschil $\Delta T_0/T_{10}$, dat 0,2% bedraagt (5 K/2600 K), kan dus worden veroorzaakt door een plaatselijke vermindering $\Delta r/r$ van de draaddiameter met slechts 0,26%, een plaatselijke vermeerdering $\Delta\varrho/\varrho$ van de specifieke weerstand met 0,8% of een plaatselijke vermindering $\Delta\varepsilon/\varepsilon$ met 0,8%. Dit is een allesbehalve volledige opsomming van mogelijke fouten. Zo kan er bij een gespiraliseerde draad sprake zijn van afwijkingen in de spoed. De specifieke weerstand kan beïnvloed zijn door korrelgrenzen, kleine holten, toevoegingen, enz. In feite kunnen heel wat verschillende typen fouten de vermelde grootheden beïnvloeden. Het is doorgaans onmogelijk te bepalen wat de invloed van al deze verschillende typen fouten op de plaatselijke energiebalans is. Men zou de factoren in beeld kunnen brengen door langs de as die de levens-

[5] O. Kubaschewski, E. Ll. Evans en C. B. Alcock, Metallurgical thermochemistry, 4e druk, Pergamon Press, Oxford 1967.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

duur voorstelt, de verschillende typen fouten met de corresponderende levensduur uit te zetten:

$\frac{\Delta r}{r}$	$\frac{\Delta \varepsilon}{\varepsilon}$	$\frac{\Delta \varrho}{\varrho}$	⟵ grootte van fout
τ_1	τ_2	τ_3	⟶ levensduur

De grootste fout is verantwoordelijk voor de waargenomen levensduur en is karakteristiek voor de technische norm. Sluit men deze fout uit, dan stelt de op één na grootste fout een grens aan de levensduur, die dan groter is. Het verschil in grootte tussen aangrenzende fouten is ook een factor van belang. Met andere woorden: wordt de levensduur wel bepaald door een fout van één bepaald type? Bij de proeven ter bepaling van het fatale gewichtsverlies werd een spreiding van $\pm$ 2% waargenomen, overeenkomend met een waarde van $\Delta(\Delta T_0)$ van $\pm$ 1,5 K. Bij een zo kleine afwijking mag men aannemen dat slechts één type fout aansprakelijk is voor de waargenomen levensduur.

Bij de afleiding van het model van de hete plek hebben we aangenomen dat de axiale warmtegeleiding langs de draad geen invloed heeft op de temperatuur T_2 van de plek. De vermelde waarden van de fouten moeten dus als ondergrens worden beschouwd. Uit een meer exacte analyse van het probleem blijkt dat bij een brandtemperatuur T_{10} van 2600 K en een draaddiameter van 50 μm de benadering van het eenvoudige model alleen geldig is voor fouten die zich uitstrekken over ten minste $60r_1$, dat is 3 mm. Bij fouten van geringere lengte moet de fout zelf groter zijn om het temperatuurverschil ΔT_0 te kunnen handhaven. Zo zal bij een fout ter lengte van slechts 5 μm de diameter van de draad 24% kleiner moeten zijn om een temperatuurverschil van 5 K te kunnen handhaven. Dezelfde waarde geldt voor andere fouten. Fouten van dergelijke grootte komen in de praktijk nooit voor. Van individuele fouten van zo geringe lengte kan dus gevoeglijk worden aangenomen dat deze geen grens stellen aan de levensduur. Het wordt echter anders als er langs de draad een groot aantal dergelijke fouten aanwezig zijn. Als deze dicht bijeen liggen, kan een van de essentiële grootheden, gemiddeld over een relatief lang stuk draad, zo sterk veranderen dat de volgens het eenvoudige model te verwachten temperatuurverhoging optreedt.

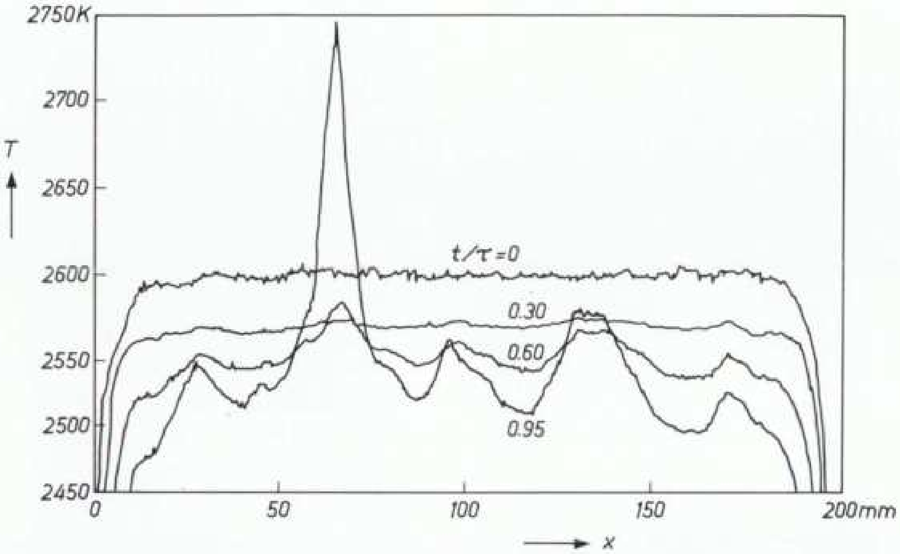


Fig. 3. Temperatuurverdeling langs een op de gebruikelijke wijze vervaardigde wolfraamdraad tijdens verschillende stadia van de levensduur. De tijd t gedurende welke de lamp heeft gebrand, is uitgedrukt als een fractie van de levensduur τ.

Vergelijking met experimentele resultaten

Het model van de hete plek berust op de veronderstelling dat de fout die tot doorbranden zal leiden, reeds aan het begin van de levensduur aanwezig is. Deze veronderstelling werd experimenteel onderzocht. Bovendien werd getracht de fout die de levensduur bepaalt, te identificeren. We hebben daartoe een elektronische micropyrometer ontwikkeld waarmee de helderheid van een rond gebied van het draadoppervlak met een diameter van 10 μm gemeten kan worden. Nemen we aan dat de plaatselijke specifieke emissie constant is, dan zal het signaal evenredig zijn met de ware temperatuur van de draad. De gevoeligheid van het instrument bedroeg ongeveer 0,5 K. Met behulp van speciaal ontwikkelde meetapparatuur konden verder verschillen in de diameter van 0,1 μm worden waargenomen. *Fig. 3* geeft de temperatuurverdeling weer, gemeten langs een op de gebruikelijke wijze gefabriceerde draad in verschillende stadia van zijn levensduur. Kort voor het doorbranden blijkt de temperatuur plaatselijk sterk op

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

te lopen. In *fig. 4* wordt de temperatuurverdeling van dezelfde draad vergeleken met de plaatselijke variatie van de draaddiameter. De opmerkelijke overeenkomst tussen beide krommen bevestigt de juistheid van de veronderstelling in het model, dat de temperatuur van de hete plek aan het einde van de levensduur uitsluitend wordt beheerst door de van de temperatuur afhankelijke verdampingsverliezen.

Uit fig. 3 blijkt dat de uiteindelijke temperatuurverdeling pas na het verstrijken van ongeveer 50% van de levensduur tot uiting komt. Op het eerste gezicht schijnt dit in tegenspraak te zijn met het model. Bovendien lijken de aanvankelijke temperatuurverschillen groter te zijn dan was voorzien. Beide verschijnselen kunnen worden verklaard door het feit dat de specifieke emissie van de draad niet overal gelijk is. De verschillen moeten toegeschreven worden aan meervoudige reflecties, veroorzaakt door de oppervlakteruwheid, en resulteren in een schijnbare verhoging van de temperatuur. De invloed van dit verschijnsel wordt geaccentueerd door de geringe afmetingen van het waargenomen gebied. De verdampingssnelheid wordt evenwel niet bepaald door de schijnbare, maar door de werkelijke temperatuur van de draad. Ten einde de vertroebelende invloed van de oppervlakteruwheid uit te sluiten, werden een aantal uitzonderlijk gladde draden vervaardigd. In *fig. 5* ziet men naast elkaar de doorsnede van een gebruikelijke en van een voor de proeven gepolijste draad. In *fig. 6* is de temperatuurverdeling langs een gepolijste draad uitgezet. Zoals een vergelijking met fig. 3 laat zien, zijn de aanvankelijke temperatuurverschillen veel kleiner, en bovendien toont de kromme $t/\tau = 0$ al op welke plaats de temperatuur het hoogst is en de hete plek ontstaat. Dit is een indicatie dat men de ware temperatuurverdeling van een draad zeer wel kan meten, mits men de invloed van de oppervlakteruwheid weet te verminderen. Door vergroting van het gebied waarvan het pyrometersignaal wordt afgeleid, b.v. van 5 µm tot een draadoppervlak ter lengte van $12\,r_1$, kan men de invloed van plaatselijke oppervlakteruwheden nagenoeg uitsluiten; het signaal komt dan goed overeen met de werkelijke temperatuur. In *fig. 7* wordt de temperatuurverdeling van een ongepolijste draad, gemeten volgens deze methode, vergeleken met die van een gepolijste draad aan het begin van de levensduur. Door het polijsten blijken de aanvankelijke temperatuurverschillen sterk verminderd te zijn. In de onderste krommen van deze figuur zijn de overeenkomstige diameterprofielen uitgezet, die duidelijk de vermindering van de oppervlakteruwheid laten zien. De verwachting is dan ook gewettigd dat een gepolijste draad met zijn geringere aanvankelijke temperatuurverschillen een langere levensduur zal hebben en dat het fatale gewichtsverlies ervan groter zal zijn. Uit proeven blijkt het fatale ge-

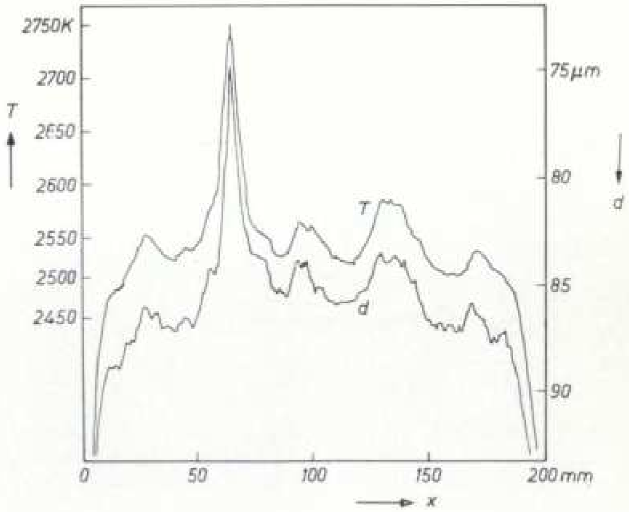


Fig. 4. Vergelijking van de temperatuur (T) met de diameter (d), tegen het einde van de levensduur, beide als functie van de afstand (x).

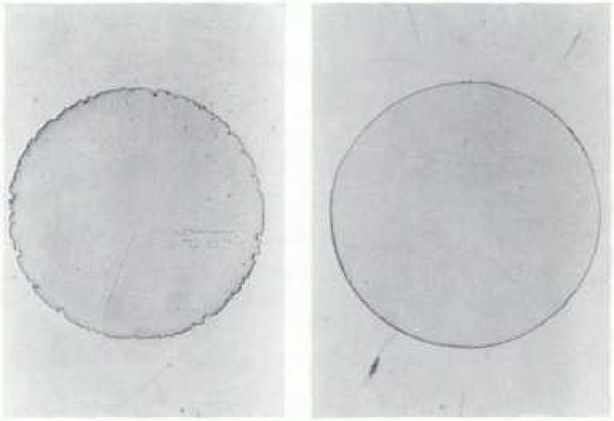


Fig. 5. Microfoto van de doorsnede van een op de gebruikelijke wijze vervaardigde wolfraamdraad (*links*, diameter 118 µm) en van een gepolijste wolfraamdraad (*rechts*, diameter 105 µm).

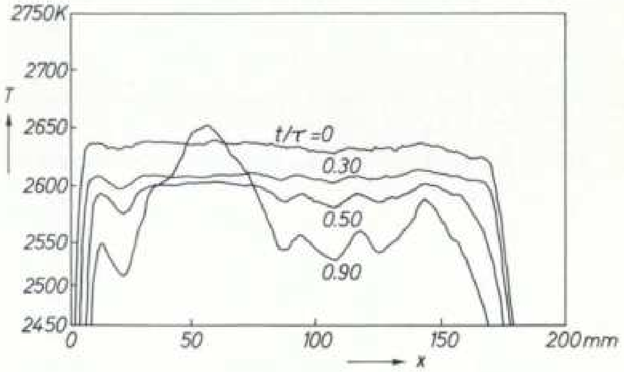


Fig. 6. Als fig. 3, nu echter voor een gepolijste draad. De aanvankelijke temperatuurverschillen langs de draad zijn veel kleiner en de plaats van de hete plek kan al aan het begin van de levensduur ($t/\tau = 0$) worden waargenomen.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

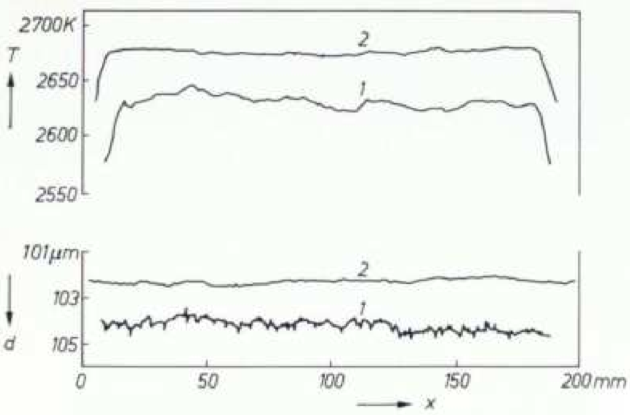


Fig. 7. Vergelijking van de temperatuur T en de diameter d, beide als functie van de afstand x; de krommen *1* gelden voor een op de gebruikelijke wijze vervaardigde wolfraamdraad, de krommen *2* voor een gepolijste.

wichtsverlies $\Delta W/W_0$ van een gewone draad ca. 18% te zijn, terwijl dat van gepolijste draden niet minder dan ca. 32% bedraagt. Het is derhalve mogelijk $\Delta W/W_0$, en dus ook $\Delta\tau/\tau$, met ongeveer 70% te verhogen, hetgeen een aanmerkelijke kwaliteitsverbetering van gloeilampen betekent.

Uit deze proeven mag men concluderen dat de levensduur van op gebruikelijke wijze vervaardigde draden wordt beheerst door de oppervlakteruwheid, die de draaddiameter en specifieke emissie bepaalt. Dit kan worden afgeleid uit het feit dat het polijsten uitsluitend het oppervlak aantast, en niet de kristalstructuur, de concentratie van toevoegingen of andere inwendige eigenschappen.

Levensduur en gewichtsverlies van een gloeispiraal in vacuüm

Door een draad te spiraliseren vermindert men de vrije oppervlakte ervan, d.w.z. het gebied dat voor de verliezen aan massa en energie aansprakelijk is. Het binnenste oppervlak van een spiraal mag men buiten beschouwing laten in de vergelijkingen voor het massa- en energieverlies, omdat het zich ongeveer gedraagt als een Knudsen-cel en als een zwarte straler. Voor gloeispiralen zoals ze in vacuümlampen worden toegepast, beslaat het vrije oppervlak $S_{\rm spiraal}$ ongeveer 50% van het overeenkomstige draadoppervlak $S_{\rm draad}$.

Men kan aantonen dat de voortschrijdende vermindering van de draaddiameter van een gespiraliseerde draad ten opzichte van een rechte wordt verminderd met een factor $S_{\rm spiraal}/S_{\rm draad}$, die ongeveer 0,5 bedraagt. Dit kan men ook interpreteren door te stellen dat de verdampingssnelheid van een spiraal ten opzichte van die van een rechte draad ongeveer gehalveerd is. Dit is in overeenstemming met proeven van Fonda [6]. In termen van het model uitgedrukt, wordt hierdoor het verloop van de temperatuur T_2 van de hete plek met de tijd en dat van de temperatuur T_1 van de foutloze draad (zie fig. 2) in de verhouding $S_{\rm spiraal}/S_{\rm draad}$ uitgerekt. Dit betekent dat de levensduur in de verhouding $S_{\rm draad}/S_{\rm spiraal}$ wordt verlengd.

Het fatale gewichtsverlies van een spiraal kan men vergelijken met dat van een rechte draad door gebruik te maken van de definitievergelijking

$$\left|\frac{\Delta W}{W_0}\right|_{\rm spiraal} = \frac{1}{W_0}\int_0^{\tau_{\rm spiraal}} S_{\rm spiraal}\, m\{T_1(t)\}\,{\rm d}t = \frac{1}{W_0}\int_0^{\frac{S_{\rm draad}}{S_{\rm spiraal}}\cdot\tau_{\rm draad}} \frac{S_{\rm spiraal}}{S_{\rm draad}}\, S_{\rm draad}\, m\{T_1(t)\}\,{\rm d}t = \left|\frac{\Delta W}{W_0}\right|_{\rm draad}. \qquad (7)$$

Deze uitkomst is in overeenstemming met de proeven van Fonda [6]. Als men mede rekening houdt met de vermindering van de verdampingsverliezen door de steundraden, kan worden aangetoond dat deze vergelijking ook van toepassing is op de in de vorige hoofdstukken vermelde resultaten.

Van het feit dat het fatale gewichtsverlies voor de gespiraliseerde en de rechte draad identiek moet zijn, zou men gebruik kunnen maken om eventuele bijkomstige fouten, veroorzaakt door het spiraliseren, te onderzoeken. Een fout die een gevolg is van deze bewerking, zou immers verraden moeten worden door een vermindering van het fatale gewichtsverlies.

Het doorbranden van lampen met gasvulling

Levensduur en gewichtsverlies van een rechte draad in een edelgas

Het gedrag van een rechte draad in een edelgasatmosfeer kunnen we in grote lijnen op dezelfde wijze beschouwen als dat van een draad in vacuüm. We zullen nu het model van de hete plek echter moeten wijzigen om enerzijds rekening te houden met de verandering in de energiebalans die wordt veroorzaakt door de warmteafgifte aan het gas, en anderzijds met de invloed van het gas op de verdampingssnelheid van het wolfraam. Bij een draadtemperatuur van 2800 K is de bijdrage van de warmteverliezen van het gas van de orde van 27%, hetgeen in dit voorbeeld een extra-stroom van 19% vereist voor het bereiken van dezelfde brandtemperatuur. Het verschil tussen vrije verdamping en door diffusie beheerst massatransport heeft overigens veel ingrijpender gevolgen. Dit massatransport M kan als volgt worden geformuleerd:

$$\boldsymbol{M} = -D\,{\rm grad}\,n + n\boldsymbol{v}, \qquad (8)$$

waarin D de diffusiecoëfficiënt voorstelt van wolfraam in het edelgas, n de wolfraamconcentratie en v de con-

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

vectiesnelheid van het edelgas. Voor de volledige oplossing van deze vergelijking moet de snelheidsverdeling bekend zijn, hetgeen doorgaans niet het geval is. Fonda [7] en Elenbaas [8] hebben een benaderende oplossing van dit probleem voorgesteld. Deze berust op de hypothese dat de gloeidraad is omgeven door een stationaire gashuid, de zg. Langmuir-laag, zodat het probleem kan worden behandeld als een geval van concentratiediffusie, waarbij de grensvoorwaarden worden gegeven door de verzadigingsconcentratie van het wolfraam aan het draadoppervlak $n(T_1)$, terwijl n aan de buitenkant van de gashuid nul is. Bij gebruikmaking van deze benadering krijgen we de volgende oplossing:

$$M = 2\pi l\, D \frac{n(T_1)}{\ln r_L/r_1}, \qquad (9)$$

waarin l de lengte van de draad voorstelt en r_L de straal van de Langmuir-laag. De verhouding g van het massaverlies van wolfraam in een edelgas tot dat in vacuüm wordt gegeven door:

$$g = \frac{M}{m} = \frac{cD}{T_1^{\frac{1}{2}} r_1 \ln r_L/r_1}. \qquad (10)$$

Deze verhouding g hangt slechts in geringe mate af van de temperatuur van de draad; voor het geval dat $r_1 = 100$ µm, de argondruk 1 atm en de draadtemperatuur 3000 K bedraagt, is g ongeveer 1/100.

Substitueren we deze uitkomst in het model, dan wordt de tijdschaal voor de temperaturen $T_1(t)$ en $T_2(t)$ met de factor $1/g$ uitgerekt. De levensduur neemt in dezelfde mate toe, zodat we krijgen:

$$\tau_{gas} = \tau_{vac}/g. \qquad (11)$$

Grijpen we terug naar vergelijking (4) en vullen we de gewijzigde grootheden m en τ in, dan blijkt het fatale gewichtsverlies niet gewijzigd te zijn.

Deze gevolgtrekking is in overeenstemming met de experimentele resultaten. De levensduur τ wordt met de overeenkomstige factor $1/g$ verlengd, zoals *fig. 8* aantoont. Door Fonda [7] werden soortgelijke uitkomsten gevonden. Voor rechte draden bedroeg het fatale gewichtsverlies gemiddeld 15%, hetgeen slechts weinig verschilt van de waarde van 18% die voor vacuümlampen werd gevonden.

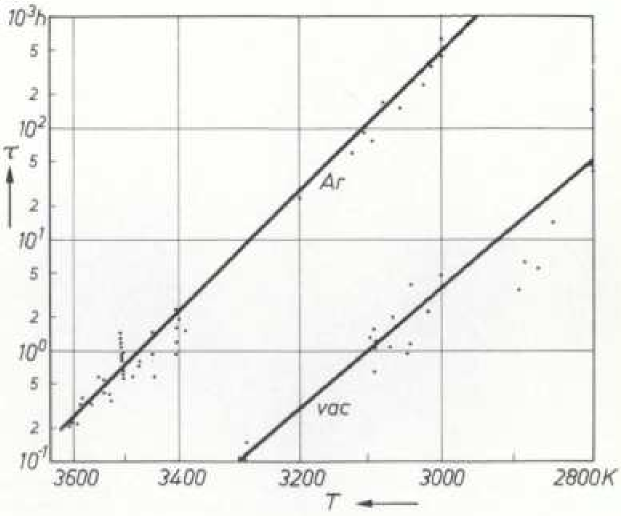


Fig. 8. Levensduur τ van rechte wolfraamdraden in vacuüm (*vac*) en in een edelgas (*Ar*, 600 torr) als functie van de brandtemperatuur T. Voor de met gas gevulde lamp is τ ongeveer een factor 100 groter, overeenkomend met de verhouding van de massaverliezen; het „fatale gewichtsverlies" is in beide gevallen gelijk.

Levensduur en fataal gewichtsverlies van een gespiraliseerde draad

Een vergelijking met metingen aan gespiraliseerde draden in een gas en in vacuüm brengt enkele interessante feiten aan het licht. Zoals te verwachten is, leidt een edelgasatmosfeer tot een langere levensduur. Het fatale gewichtsverlies daalt echter van 15% tot 2%. Nu bleek uit de voorgaande beschouwingen over het model van een rechte draad, dat de levensduur en het fatale gewichtsverlies door hetzelfde fysische proces worden beheerst, nl. door radiale diffusie van het wolfraam. Als het model van de hete plek nog geldig zou zijn, dan zou het fatale gewichtsverlies gelijk moeten blijven. De geconstateerde discrepantie vereist dan ook een herziening van het model.

Bekende empirisch gevonden feiten, zoals het verband tussen de levensduur en de temperatuur van de spiraal, de druk van de gasvulling en de aard ervan, wijzen erop dat de diffusie van het wolfraam ook niet het proces is dat de levensduur beheerst. We moeten dus wel aannemen dat er naast de radiale diffusie nog een ander diffusieproces optreedt, dat overheerst.

Bij het vorsen naar een ander diffusiemechanisme stuit men op de alleszins aannemelijke mogelijkheid dat verdampt wolfraam binnen de gloeispiraal zelf wordt getransporteerd [3]. Een dergelijk proces kan worden veroorzaakt door temperatuurverschillen tussen diverse delen van de spiraal. *Fig. 9* geeft hiervan een voorbeeld. Fig. 9*a* toont een winding van een dubbel gespiraliseerde gloeidraad, waarvan fig. 9*b* de doorsnede voorstelt, tegen het einde van de levensduur. Er blijkt een massatransport te hebben plaatsgehad van de binnenste delen van de individuele spiraalwindingen naar de buitenste delen ervan. Een kwantitatieve analyse van het diffusieproces in een zo gecompliceerde geometrische configuratie is verre van eenvoudig. Het probleem wordt echter wat overzichtelijker in het geval

[6] G. R. Fonda en A. A. Vernon, J. Opt. Soc. Amer. **22**, 223, 1932.
[7] G. R. Fonda, Phys. Rev. **21**, 343, 1923 en **31**, 260, 1928.
[8] W. Elenbaas, Philips Res. Repts. **18**, 147, 1963.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

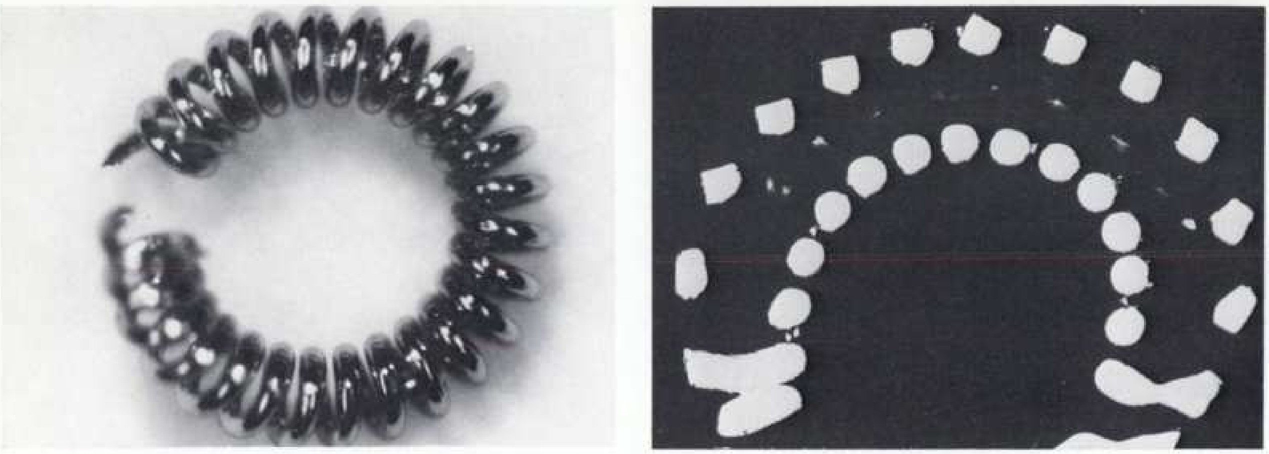


Fig. 9. *Links:* voorbeeld van een „dubbel gespiraliseerde" gloeidraad. *Rechts:* microfoto van een dergelijke draad nadat deze is doorgebrand. De diameters van de buitenste delen van de windingen zijn groter geworden dan die van de oorspronkelijke draad, hetgeen wijst op een massatransport van de binnenste delen naar de buitenste.

van een laagspanningslamp. De grote invloed van de axiale warmtegeleiding geeft dan aanleiding tot een sterke daling van de temperatuur in de axiale richting, met als gevolg een axiale gradiënt van de wolfraamconcentratie [4]. Deze veroorzaakt een aanzienlijk axiaal massatransport M_{ax}, dat onder bepaalde omstandigheden het radiale transport zal overtreffen.

Men kan het wolfraamtransport binnen een enkele spiraal berekenen, door te stellen dat tussen aangrenzende windingen ervan een temperatuurverschil ΔT bestaat, zie *fig. 10*. Dit verschil veroorzaakt een concentratieverschil, gegeven door $n(T) = n_0 T^{\gamma}$ en $\Delta n = \gamma n(T) \Delta T/T$. De axiale diffusie tussen aangrenzende windingen met temperaturen T en $T - \Delta T$ bedraagt dan:

$$M_{ax} = -2\pi r_2\, D \int \text{grad}\, n\, df =$$

$$= \frac{2\pi r_2\, D\, \pi \Delta n}{\ln\left[\dfrac{S}{r_1} + \sqrt{\left(\dfrac{S}{r_1}\right)^2 - 1}\right]}, \qquad (12)$$

waarin $2S$ de spoed en r_2 de straal van de spiraal voorstelt.

Gebruiken we de notatie volgens fig. 10, dan krijgen we voor het axiale massatransport:

$$M_{ax} = \frac{2\pi r_2\, D\, \pi \Delta n}{\ln 2{,}6} \approx 2\pi^2 r_2\, D\, \gamma\, n(T) \frac{\Delta T}{T}. \quad (13)$$

Dit transport moet worden vergeleken met het radiale transport van één winding. Men kan dit berekenen door aan te nemen dat de wolfraamconcentratie op een cilindrische mantel rondom de spiraal ter lengte van $3r_1$ bij benadering constant is en aan de Langmuir-grens gelijk nul is.

$$M_{rad} = D\,.\,2\pi\,.\,3r_1 \frac{n(T)}{\ln r_L/r_2}. \qquad (14)$$

Volgens (13) en (14) vindt men voor de verhouding van het axiale tot het radiale massatransport bij benadering:

$$\frac{M_{ax}}{M_{rad}} = \frac{\pi r_2\, \gamma \ln r_L/r_2}{3r_1} \frac{\Delta T}{T}. \qquad (15)$$

De voorwaarde die ons in het bijzonder interesseert, is $M_{ax}/M_{rad} \geqslant 1$. Deze definieert immers het temperatuurverschil waarbij het axiale transport gelijk is aan het radiale transport, of dit overtreft:

$$\Delta T \geqslant \frac{3Tr_1}{\pi\, \gamma\, r_2 \ln r_L/r_2}. \qquad (16)$$

We vinden voor $T = 3000$ K, $r_1/r_2 = 0{,}1$ en $r_L/r_2 = 5$ een temperatuurverschil ΔT van ca. 5 K.

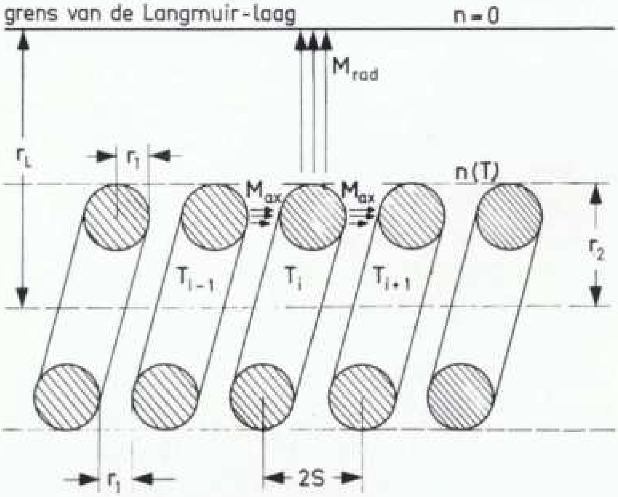


Fig. 10. Schematische voorstelling van het wolfraamtransport in een enkelvoudige gloeispiraal ten gevolge van radiale (M_{rad}) en axiale (M_{ax}) diffusie. De indices . . ., $i - 1$, i, $i + 1$, . . . geven het nummer van de winding.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Dit is geringer dan het temperatuurverschil tussen aangrenzende windingen in laagspanningslampen. We mogen dan ook concluderen dat de levensduur van deze lampen wordt bepaald door het *axiale* massatransport. Een typisch voorbeeld van een lamp waarin het axiale transport overheerst, is een projectielamp (12 V, 100 W). In *fig. 11a* is de temperatuurverdeling van de gloeispiraal hiervan uitgezet. Door toepassing van verg. (13) en vereenvoudiging van de notatie, vinden we dat het axiale transport over de lengte x van de spiraal varieert volgens:

$$M_{ax} = c\, n(T)\, \Delta T/T = c\, T^{34}(x)\, \Delta T(x), \qquad (17)$$

aangezien γ bij de betrokken temperatuur (3200 K) gelijk is aan 35.

Uit deze vergelijking volgt dat bij de temperatuurverdeling volgens fig. 11*a* in het midden en aan de uiteinden minima in het axiale transport ontstaan. Dit is een gevolg van de lage waarde van ΔT in het midden en van de aanmerkelijk lagere temperatuur van de uiterste windingen.

Ter weerszijden van het centrale minimum moeten maxima aanwezig zijn. Men kan dus verwachten dat bij een overwegend axiaal massatransport de gloeidraad zal doorbranden ter plaatse van een maximum in M_{ax}, dat in dit geval niet samenvalt met de plaats waar de temperatuur het hoogst is. Men bedenke evenwel dat het verloop van de temperatuur van een winding met de tijd geen functie is van M_{ax}, maar van ΔM_{ax}, het netto axiale transport; dit is voor de winding met rangnummer i gelijk aan het verschil tussen het massatransport van winding $i-1$ naar winding i en het transport van de laatstgenoemde winding naar winding $i+1$ (zie fig. 10):

$$\Delta M_{ax,i} = C\left[T^{\gamma}_{i-1}\,\Delta T_{i-1,i} - T_i^{\gamma}\,\Delta T_{i,i+1}\right], \qquad (18)$$

waarin $\Delta T_{i-1,i} = T_{i-1} - T_i$ en $\Delta T_{i,i+1} = T_i - T_{i+1}$. Volgens de berekening neemt voor een temperatuurverdeling volgens fig. 11*a* het netto axiale transport de vorm aan volgens fig. 11*b*. De negatieve waarden van ΔM_{ax} stellen het massa*verlies* van een winding voor. Fig. 11*c* is een foto van dezelfde gloeidraad nadat deze is doorgebrand. Uit een vergelijking met fig. 11*b* blijkt dat de plaats waar de draad is doorgebrand, samenvalt met een maximum van het massaverlies. De karakteristieke eigenschappen van fig. 11*c* zijn typerend voor een axiaal massatransport: de winding met het grootste massaverlies is niet alleen dunner geworden, maar door thermisch etsen tevens gladder, terwijl de aangrenzende windingen verschijnselen vertonen van dendritische groei.

Fig. 12 laat nóg een voorbeeld zien van dit verschijnsel; het betreft hier de gloeidraad van een doorgebrande autolamp. De gloeidraad blijkt niet in het midden door te branden, waar de temperatuur het hoogst is, maar op een plaats met een lagere temperatuur, waar het netto axiale massatransport maximaal is. Zoals bij een symmetrische warmteverdeling kan worden verwacht, ontstaat aan beide zijden van het midden een plek waar de gloeidraad wordt aangetast. Dit is een bewijs te meer dat axiale verdamping de levensduur beheerst. Zowel de levensduur als de plaats waar de gloeidraad zal doorbranden, kunnen dus wor-

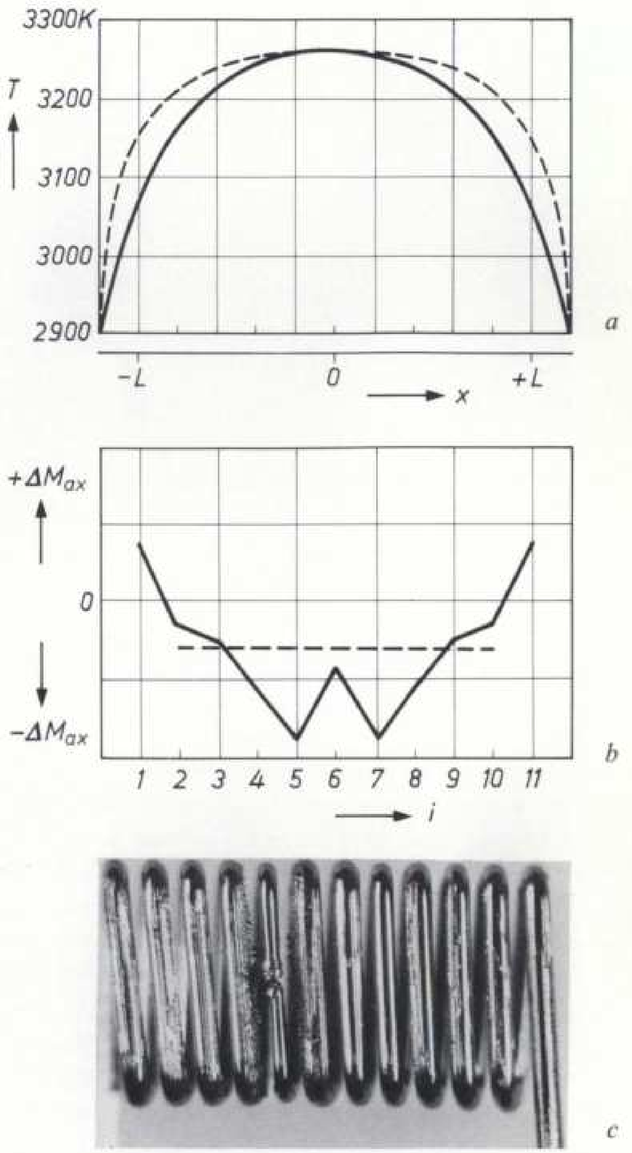


Fig. 11. *a)* Temperatuurverdeling langs een gespiraliseerde gloeidraad van een projectielamp (12 V, 100 W). De letter i stelt weer het rangnummer van de windingen voor. De getrokken kromme komt overeen met het gemeten resultaat; de onderbroken kromme is in overeenstemming met verg. (19) en verzekert een constant axiaal massatransport over de gehele lengte. *b)* Netto axiaal massatransport volgens verg. (18), berekend uit de gemeten temperatuurverdeling (getrokken kromme) en uit de volgens verg. (19) berekende temperatuurverdeling (onderbroken lijn). *c)* Foto van de doorgebrande gloeidraad; de aangetaste plekken ter plaatse van de maxima van het axiale massaverlies kunnen duidelijk worden onderscheiden.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

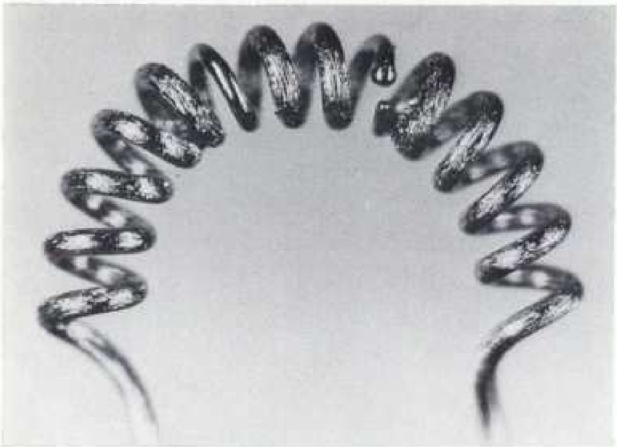


Fig. 12. Gloeidraad van een doorgebrande autolamp (6 V, 40 W) met de typische symmetrische ligging van de verzwakte plekken ter weerszijden van het midden.

den voorspeld door een analyse van de warmteverdeling, die op haar beurt een functie is van de geometrische configuratie van de gloeidraad.

De bovenstaande beschouwingen zijn eveneens van toepassing op laagspanningshalogeenlampen. Deze onderscheiden zich in zoverre van normale met gas gevulde lampen, dat het wolfraam niet naar de ballon, maar van hetere naar koudere delen van de gloeidraad wordt getransporteerd. In het gebied van de gloeidraad waar de temperatuur hoog is, wordt het wolfraamtransport echter door atomaire diffusie beheerst, evenals in lampen die uitsluitend met edelgas zijn gevuld. Dit wordt verklaard door de geringe stabiliteit van jodium en broomverbindingen van wolfraam. In halogeenlampen is het dus eveneens de axiale wolfraamdiffusie die het proces van het doorbranden beheerst. Dat bij halogeenlampen een langere levenscuur wordt waargenomen dan bij overeenkomstige alleen met edelgas gevulde lampen, moet worden toegeschreven aan de hogere druk van het edelgas (p is ca. 1 atm voor normale lampen, tegen 5 à 10 atm voor halogeenlampen); hierdoor is de diffusiecoëfficiënt D van het wolfraam kleiner — deze is omgekeerd evenredig met p — en het door axiale diffusie veroorzaakte transport dus geringer. Dat het axiale transport bij halogeenlampen inderdaad aansprakelijk is voor het doorbranden, wordt geïllustreerd door *fig. 13*, die de gloeidraad van een halogeenlamp (type H3) laat zien tegen het einde van zijn levensduur. De typische kenmerken van diffusie van de ene winding van de gloeispiraal naar de andere treden ook hier duidelijk aan het licht.

De axiale diffusie bij laagspanningslampen

In het voorgaande hoofdstuk werd aangetoond dat de levensduur van laagspanningslampen wordt beheerst door het axiale massatransport en dat dit transport geheel wordt bepaald door de axiale temperatuurverdeling. Het ligt dan ook voor de hand te veronderstellen dat men de levensduur kan verlengen door de temperatuurverdeling langs de as van de gloeispiraal zo te wijzigen dat de maxima in het axiale transport (zie fig. 11*b*) worden verlaagd. Men kan de in dit opzicht gunstigste temperatuurverdeling afleiden door als voorwaarde te stellen dat ΔM langs de gloeispiraal constant is. Deze voorwaarde kan men als volgt formuleren:

$$\partial M/\partial x = \text{constant},$$

waarin M het totale massatransport per winding voorstelt; in het algemeen bestaat M uit een axiale en een radiale bijdrage. De oplossing van deze differentiaalvergelijking leidt tot volgende temperatuurverdeling:

$$T(x) = T_{max}\left[1 - \left\{1 - \left(\frac{T_{min}}{T_{max}}\right)^{\gamma}\right\} \times \frac{\exp(|x|/r_L) - (1 + |x|/r_L)}{\exp(L/r_L) - (1 + L/r_L)}\right]^{1/\gamma}, \qquad (19)$$

waarin T_{max} de maximale temperatuur in het midden van de gloeispiraal voorstelt ($x = 0$), T_{min} de minimale temperatuur aan de uiteinden ($x = \pm L$), r_L de straal van het Langmuir-gebied en L de halve lengte van de gloeispiraal. Uitgaande van dezelfde waarden van T_{max} en T_{min} als die van de gloeidraad (12 V, 100 W) waarvan de temperatuurverdeling in fig. 11*a* is uitgezet, verkrijgen we volgens (19) de ter vergelijking eveneens in deze grafiek getekende kromme (onderbroken lijn). De aldus berekende waarden van $\Delta M(x)$, overeenkomende met deze temperatuurprofielen, zijn in fig. 11*b* uitgezet. Voor de ideale temperatuurverdeling blijkt $\Delta M(x)$ constant te zijn en lager dan de aan de gloeidraad gemeten waarden. Men zou de waarde nog verder kunnen verlagen door het temperatuurverschil tussen het midden en de uiteinden van de gloeidraad te verminderen. Dit werd bevestigd door laboratoriumproeven, die aantoonden dat men op deze wijze de levensduur ongeveer kan verdubbelen. Op dit onderwerp wordt in een afzonderlijk artikel ingegaan.

* * *

Tot voor kort was het inzicht in het mechanisme dat zich in gloeilampen afspeelt, voornamelijk gebaseerd op de kennis van een aantal welbekende, empirisch vastgestelde feiten. Deze verschaften echter geen consistent beeld van de fysische processen die zich afspelen. Door de invoering van een model, namelijk dat van de hete plek, kunnen de meeste van deze feiten in verband worden gebracht met een beperkt aantal fysische grootheden. Een van de belangrijkste gevolgtrekkingen waartoe dit model leidt, is dat de levensduur van een gloei-

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

draad uitsluitend wordt bepaald door de beginvoorwaarden ervan, te weten de temperatuur van de draad en die van de hete plek. Het model wordt hier voor het eerst toegepast om een analytische formule voor de levensduur af te leiden. Deze blijkt niet alleen af te hangen van de verdampingssnelheid, maar ook van de grootte van de aanvankelijke hete plek. Bovendien verschaft de formule een fysisch inzicht in het fatale gewichtsverlies, een grootheid die tot nu toe als maat voor de kwaliteit werd gebruikt zonder dat het belang ervan ten volle werd ingezien.

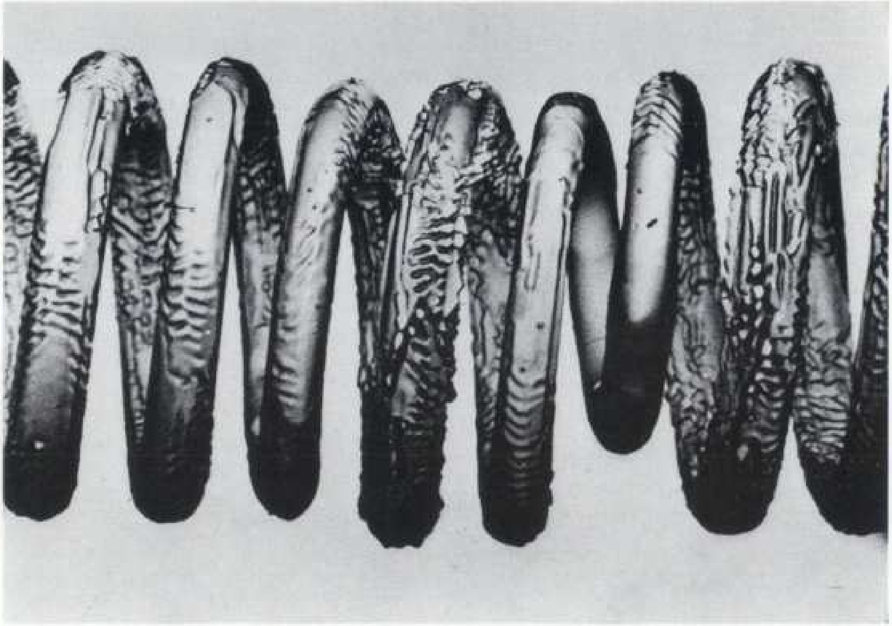


Fig. 13. De gloeidraad van een halogeenlamp voor auto's (H3, 12 V, 55 W) tegen het einde van zijn levensduur; ook hier zien wij de typische kenmerken van axiale diffusie tussen de windingen onderling.

Voor het geval van een rechte draad blijkt uit het model dat een aanvankelijk plaatselijk temperatuurverschil van ongeveer 5 K de levensduur al kan bepalen. Een dergelijk verschil kan te wijten zijn aan oppervlakteruwheden, die aanleiding geven tot verschillen in de specifieke emissie en de draaddiameter.

Een nauwgezet onderzoek naar de temperatuurverdeling langs de draad brengt aan het licht dat er verscheidene hete plekken kunnen voorkomen. Niettemin is het model vrijwel in volmaakte overeenstemming met de experimenten. De resultaten worden nauwelijks gewijzigd door een gecompliceerder model in te voeren, dat rekening houdt met een meer realistisch verloop van de temperatuurverdeling met de tijd.

Past men dit model toe op een gespiraliseerde draad in een lamp met gasvulling, dan blijkt het niet in overeenstemming te zijn met de experimentele resultaten. Dit leidde ertoe een model in te voeren dat tevens rekening houdt met het transport van wolfraam binnen de gloeispiraal. Bij een laagspanningslamp bestaat dit proces uit axiale diffusie van wolfraam van de ene winding naar de andere; deze diffusie vindt haar oorzaak in axiale temperatuurgradiënten. Men kan de plaats waar de draad zal doorbranden voorspellen door uit een gegeven temperatuurprofiel het axiale massatransport te berekenen. Het model kan tevens dienst doen voor de afleiding van een temperatuurprofiel dat een minimaal axiaal massatransport en een overeenkomstig langere levensduur verzekert.

Het ligt in de verwachting dat de met dit model bereikte resultaten de ontwikkeling zullen bevorderen van de technologie van wolfraamdraad en dat dit zal leiden tot een kwaliteitsverbetering van gloeilampen.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Philips techn. T. 32, 316-329, 1971, no. 10/11/12

Verstrooiingseigenschappen van een kruiskoppeling van overmaatse golfpijpen

C. J. Bouwkamp

Inleiding

Enige jaren geleden heeft schrijver dezes de reflectie-, transmissie- en deflectie-eigenschappen van een kruiskoppeling van overmaatse golfpijpen bestudeerd voor het geval dat de koppeling wordt aangeslagen door de dominante transversaal-elektrische modus die invalt vanuit het oneindige en zich voortplant in één van de vier armen die samen de kruiskoppeling vormen. Dit onderzoek werd verricht op verzoek van dr. H. J. Butterweck, toen verbonden aan ons Laboratorium en nu hoogleraar in de Theoretische Elektrotechniek aan de Technische Hogeschool te Eindhoven. De resultaten ervan werden vastgelegd in een intern rapport dat nooit is gepubliceerd. Het lijkt ons echter gerechtvaardigd thans tct publikatie over te gaan, en wel om ten minste twee redenen. Ten eerste bleek dat onze „exacte" oplossing van het verstrooiingsprobleem kan dienen als controle van de methodiek van J. B. Keller betreffende de geometrisch-optische diffractie, wanneer deze wordt toegepast op golfgeleidersystemen, zoals wordt aangetoond door het werk van L. B. Felsen c.s. van het Brooklyn Polytechnic Institute. Ten tweede houdt onze analyse verband met problemen buiten het elektromagnetische vakgebied, zoals blijkt uit een onderzoek dat onlangs is verricht onder leiding van prof. dr. J. Boersma van de Onderafdeling Wiskunde der Technische Hogeschool te Eindhoven. Het lijkt ons juist, hier te vermelden dat Boersma enige van de vroeger aangetroffen moeilijkheden heeft kunnen uitschakelen en vragen heeft kunnen beantwoorden die waren opengebleven.

De oorspronkelijke golf, zie *fig. 1*, wordt ondersteld in te vallen vanuit het oneindige in gebied *I*. Dientengevolge ontstaat een oneindig systeem van secundaire golven in ieder van de vier armen van de kruiskoppeling. Slechts een eindig aantal van deze secundaire golven plant zich voort, de overige zijn gedempt. In het bijzonder wanneer de golfgeleider overmaats is, doen zich voortplantende modi van hogere orde voor naast de dominante modus. In gebied *I* zullen teruggekaatste golven voorkomen, in gebied *II* doorgelaten golven, en in de gebieden *III* en *IV* afgebogen golven.

Prof. dr. C. J. Bouwkamp, wetenschappelijk adviseur van Philips' Natuurkundig Laboratorium, Eindhoven.

Het doel van dit artikel is het hierboven beschreven verstrooiingsprobleem mathematisch te analyseren en betrekkingen af te leiden voor de verschillende reflectie-, transmissie- en deflectiecoëfficiënten, die dienstig zullen zijn voor de numerieke berekening van deze grootheden met behulp van een computer. Hierbij zullen numerieke gegevens beschikbaar komen ter ondersteuning van de experimentele resultaten van dr. Butterweck.

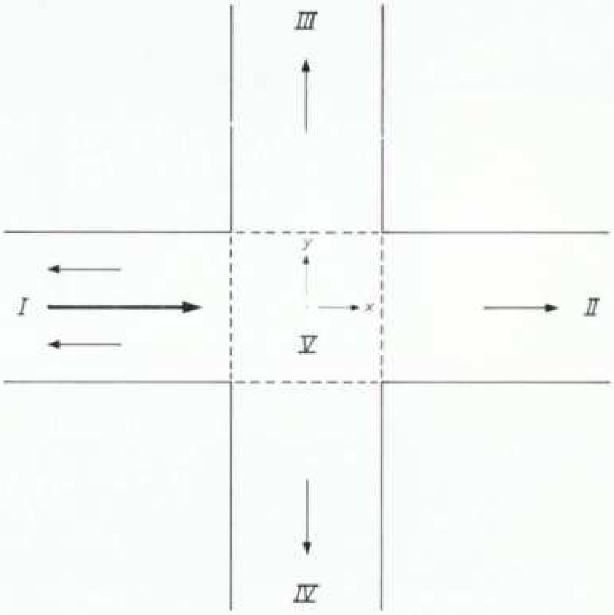


Fig. 1. Kruiskoppeling van golfpijpen. De invallende golf komt van links in gebied *I*.

Ontwikkeling in modi

We onderstellen dat de tijdsafhankelijkheid van het veld wordt gegeven door exp $(-i\omega t)$, waarin i de imaginaire eenheid aanduidt. Deze tijdfactor zal steeds worden weggelaten. Verder zij k het golfgetal, $k = \omega/c$, waarin c de lichtsnelheid voorstelt.

Laat x en y rechthoekige cartesische coördinaten in het vlak van tekening zijn, met de oorsprong in het centrum van het kruis, zoals is getekend in fig. 1, en met de z-as loodrecht op dit vlak en wijzend naar de lezer. Doordat de invallende golf (dominant) transversaal-elektrisch is, is E_z de enige component van het elektrische veld die ongelijk nul is en ligt de magnetische vector in het vlak van tekening. De vergelijkingen van

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Maxwell welke betrekking hebben op deze situatie zijn:

$$E_x = E_y = H_z = 0,$$
$$ik\,H_x = \partial E_z/\partial y,$$
$$-ik\,H_y = \partial E_z/\partial x,$$
$$\partial H_x/\partial y - \partial H_y/\partial x = ik\,E_z,$$

waarbij wij de eenheden van Gauss hebben gebruikt.

Hieruit volgt dat ons verstrooiingsprobleem in feite een scalair probleem is: het elektromagnetische veld kan worden beschreven door de scalaire golffunctie $\Phi = E_z$ die voldoet aan de tijdsonafhankelijke tweedimensionele golfvergelijking $\Delta\Phi + k^2\Phi = 0$ onder geschikte rand- en stralingscondities. Aangezien we zullen aannemen dat de wanden van de golfgeleider ideaal geleidend zijn, moet aan de wand $\Phi = 0$ gelden.

Laten we nu de mogelijke scalaire modi van een oneindige golfpijp met breedte a beschouwen, en coördinaten x',y' kiezen, zoals is weergegeven in *fig. 2*. Als n een positief geheel getal is, is de functie

$$\sin(n\pi y'/a)\exp[i(\pi x'/a)\sqrt{\{(ka/\pi)^2 - n^2\}}]$$

een oplossing van de golfvergelijking die nul is aan de

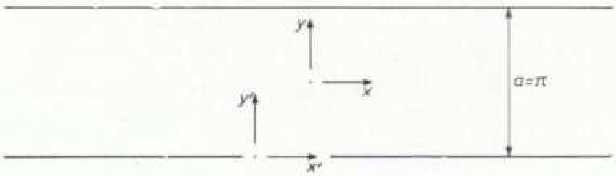


Fig. 2. De twee coördinatenstelsels.

rand ($y' = 0,a$) en een golf voorstelt die zich voortbeweegt in de positieve x'-richting. Ten einde de notatie te vereenvoudigen, kiezen we voor het gemak de dimensieloze grootheden $\pi y'/a$, $\pi x'/a$ en ka/π, en duiden deze weer aan met respectievelijk y', x', k. Met andere woorden: de breedte van de golfpijp wordt genormeerd op $a = \pi$. De bovenstaande oplossing wordt dan:

$$\sin(ny')\exp[ix'\sqrt{(k^2 - n^2)}].$$

Deze functie representeert een zich voortbewegende golf indien $k > n$ is, en een uitdempende golf als $k < n$ is. In dit verband moeten we voorzichtig zijn met de definitie van de vierkantswortel. Laat daarom $(p)^{\frac{1}{2}}$ de positieve waarde van de vierkantswortel uit het positieve getal p voorstellen. Dan definiëren we:

$$\sqrt{(k^2 - n^2)} := \begin{cases} (k^2 - n^2)^{\frac{1}{2}} \text{ als } k > n, \\ i(n^2 - k^2)^{\frac{1}{2}} \text{ als } k < n. \end{cases}$$

In het tweede geval moet de uitkomst zijn $+i(n^2 - k^2)^{\frac{1}{2}}$, ten einde dan een gedempte golf in de positieve x'-richting te verkrijgen.

Om moeilijkheden met resonantie te vermijden, eisen we uitdrukkelijk dat k geen positief geheel getal is, of anders nemen we aan dat k complex is met een klein doch positief imaginair gedeelte. Dan wordt in het algemeen $\sqrt{(k^2 - n^2)}$ zo gedefinieerd dat geldt: $0 \leqslant \arg[\sqrt{(k^2 - n^2)}] \leqslant \frac{1}{2}\pi$, met een bijbehorende analytische voortzetting naar de positief-reële k-as.

De coördinaten x en y van fig. 1 zijn het meest geschikt om volledig voordeel te trekken uit de symmetrie van de kruiskoppeling en van de invallende golf, die een even functie van y is. Een representatie van de modi is echter eenvoudiger met de coördinaten x',y' van fig. 2. Indien we de oorsprong van x',y' verplaatsen over afstanden $\frac{1}{2}\pi$, dan verkrijgen we het x,y-stelsel van fig. 2, waarbij tussen de twee stelsels de betrekkingen $x' = x + \frac{1}{2}\pi$, $y' = y + \frac{1}{2}\pi$ bestaan. In het geval van deze nieuwe coördinaten x en y zijn de eigenfuncties

$$\sin\{n(y + \tfrac{1}{2}\pi)\}\exp[i(x + \tfrac{1}{2}\pi)\sqrt{(k^2 - n^2)}],$$

of, afhankelijk van de pariteit van n,

$$\sin(\tfrac{1}{2}n\pi)\cos(ny)\exp[i(x + \tfrac{1}{2}\pi)\sqrt{(k^2 - n^2)}] \quad (n \text{ oneven})$$
$$\cos(\tfrac{1}{2}n\pi)\sin(ny)\exp[i(x + \tfrac{1}{2}\pi)\sqrt{(k^2 - n^2)}] \quad (n \text{ even}).$$

Indien we dit wensen, kunnen we de fasefactor $\exp[\frac{1}{2}\pi i\sqrt{(k^2 - n^2)}]$ weglaten of vervangen door welke andere dan ook, mits deze niet afhangt van x of y.

De dominante modus wordt verkregen voor $n = 1$:

$$\cos y \exp[i(x + \tfrac{1}{2}\pi)\sqrt{(k^2 - 1)}].$$

Wil dit een zich voortplantende golf zijn, dan moet $k > 1$ gelden.

Zoals eerder is opgemerkt, is deze dominante modus een even functie van y (zie fig. 1; voortaan zal de breedte van iedere arm gelijk aan π worden genomen). Dit houdt in dat de hele veldconfiguratie een even functie van y is. Dienovereenkomstig zullen in de gebieden *I* en *II* alleen modi met oneven n ontstaan. Anderzijds zullen de zijwaarts afgebogen modi in de gebieden *III* en *IV*, die verkregen zijn uit de modi in de gebieden *I* en *II* door het verwisselen van x en y, alle natuurlijke getallen n, zowel even als oneven, omvatten.

Met betrekking tot fig. 1 zullen wij het totale veld in ieder van de vier gebieden *I-IV* voorstellen zoals is aangegeven in *tabel I*, en wel door middel van een stel coëfficiënten A_n, B_n, C_n en D_n. Deze ontwikkeling in modi laat zien dat de oorspronkelijke golf niet voorkomt in de gebieden *III* en *IV*, en gelijk is aan $\cos y\exp[i(x + \frac{1}{2}\pi)\sqrt{(k^2 - 1)}]$ in gebied *I* èn gebied *II*. Aan het eind van gebied *I*, d.w.z. ter plaatse $x = -\frac{1}{2}\pi$, is deze functie eenvoudig $\cos y$. Dit betekent dat de amplitude van de invallende golf op één wordt genormeerd in het punt waar de golf de kruiskoppeling (gebied *V*) binnentreedt.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Tabel I. Coëfficiënten van de ontwikkeling in modi voor de vier gebieden van fig. 1.

Gebied *I* $(-\infty < x < -\frac{1}{2}\pi,\ -\frac{1}{2}\pi < y < \frac{1}{2}\pi)$:

$$\phi_I = \cos y \exp\left[i(x+\tfrac{1}{2}\pi)\sqrt{(k^2-1)}\right] + \sum_{n \text{ oneven} > 0} (B_n - D_n)\sin(\tfrac{1}{2}n\pi)\cos(ny)\exp\left[-i(x+\tfrac{1}{2}\pi)\sqrt{(k^2-n^2)}\right]$$

Gebied *II* $(\frac{1}{2}\pi < x < \infty,\ -\frac{1}{2}\pi < y < \frac{1}{2}\pi)$:

$$\phi_{II} = \cos y \exp\left[i(x+\tfrac{1}{2}\pi)\sqrt{(k^2-1)}\right] + \sum_{n \text{ oneven} > 0} (B_n + D_n)\sin(\tfrac{1}{2}n\pi)\cos(ny)\exp\left[i(x-\tfrac{1}{2}\pi)\sqrt{(k^2-n^2)}\right]$$

Gebied *III* $(-\frac{1}{2}\pi < x < \frac{1}{2}\pi,\ \frac{1}{2}\pi < y < \infty)$:

$$\phi_{III} = \sum_{n \text{ oneven} > 0} A_n \sin(\tfrac{1}{2}n\pi)\cos(nx)\exp\left[i(y-\tfrac{1}{2}\pi)\sqrt{(k^2-n^2)}\right] - \sum_{n \text{ even} > 0} C_n\cos(\tfrac{1}{2}n\pi)\sin(nx)\exp\left[i(y-\tfrac{1}{2}\pi)\sqrt{(k^2-n^2)}\right]$$

Gebied *IV* $(-\frac{1}{2}\pi \leqslant x \leqslant \frac{1}{2}\pi,\ -\infty < y \leqslant -\frac{1}{2}\pi$:

$$\phi_{IV} = \sum_{n \text{ oneven} > 0} A_n \sin(\tfrac{1}{2}n\pi)\cos(nx)\exp\left[-i(y+\tfrac{1}{2}\pi)\sqrt{(k^2-n^2)}\right] - \sum_{n \text{ even} > 0} C_n\cos(\tfrac{1}{2}n\pi)\sin(nx)\exp\left[-i(y+\tfrac{1}{2}\pi)\sqrt{(k^2-n^2)}\right]$$

De secundaire velden in de eerste vier gebieden worden ontwikkeld in oneindige reeksen van modi. In alle termen worden de fasefactoren zodanig gekozen dat zij gelijk aan één worden (fase nul) op de rand van gebied *V*. Dit is opnieuw een kwestie van normeren, maar het is zeer praktisch.

Wat wij proberen te vinden zijn de coëfficiënten A_n, B_n, D_n (n oneven) en C_n (n even); dit zijn complexe getallen. Met behulp van een mathematische analyse van het verstrooiingsprobleem moeten we kunnen vaststellen hoe ze van k afhangen.
De *reflectie*coëfficiënten voor gebied *I* zijn:

$$R_n = 0 \quad (n \text{ even}), \qquad R_n = B_n - D_n \quad (n \text{ oneven}).$$

De *transmissie*coëfficiënten voor gebied *II* zijn:

$$T_n = 0 \ (n \text{ even}), \qquad T_n = B_n + D_n \ (n \text{ oneven}, \neq 1),$$
$$T_1 = B_1 + D_1 + \exp\left[\pi i \sqrt{(k^2-1)}\right].$$

De *deflectie*coëfficiënten voor de gebieden *III* en *IV* zijn:

$$S_n = A_n \ (n \text{ oneven}) \quad \text{en} \quad S_n = -C_n \ (n \text{ even}).$$

Het is onze bedoeling deze coëfficiënten te bepalen en hiervan de amplituden en fasen in een tabel onder te brengen voor $k = 1(0.1)13$ [*] met k niet geheel, en wel tot $n = 30$.

Formulering in termen van integraalvergelijkingen

Ons verstrooiingsprobleem kan worden gesplitst in twee los van elkaar staande delen. Een ervan is een even functie van x, het andere is oneven in x. Om dit in te zien, ontbinden we de oorspronkelijke exciterende golf in twee staande golven u_1 en u_2. De ene,

$$u_1 = \exp\left[\tfrac{1}{2}\pi i \sqrt{(k^2-1)}\right]\cos y \cos\left\{x\sqrt{(k^2-1)}\right\},$$

is een even functie van x, de andere,

$$u_2 = i\exp\left[\tfrac{1}{2}\pi i \sqrt{(k^2-1)}\right]\cos y \sin\left\{x\sqrt{(k^2-1)}\right\},$$

is een oneven functie van x.

Als de kruiskoppeling wordt aangeslagen door u_1 alleen, zal het hele veld even in x zijn; als de koppeling wordt aangeslagen door u_2, zal het veld oneven in x zijn. Door optelling van de twee partiële velden verkrijgen we het veld dat ontstaat door de invallende lopende golf $u_1 + u_2$.

We zullen ons eerst bezighouden met het even probleem. Als voorbereiding lossen we het volgende randwaardeprobleem van Dirichlet voor de golfvergelijking op (zie voor de notatie *fig. 3*). Zij $f_1(x) = f_1(-x)$ een even functie van x, die is gedefinieerd in het interval

[*] In dit artikel wordt als decimaalteken de *punt* gebruikt in plaats van de komma.

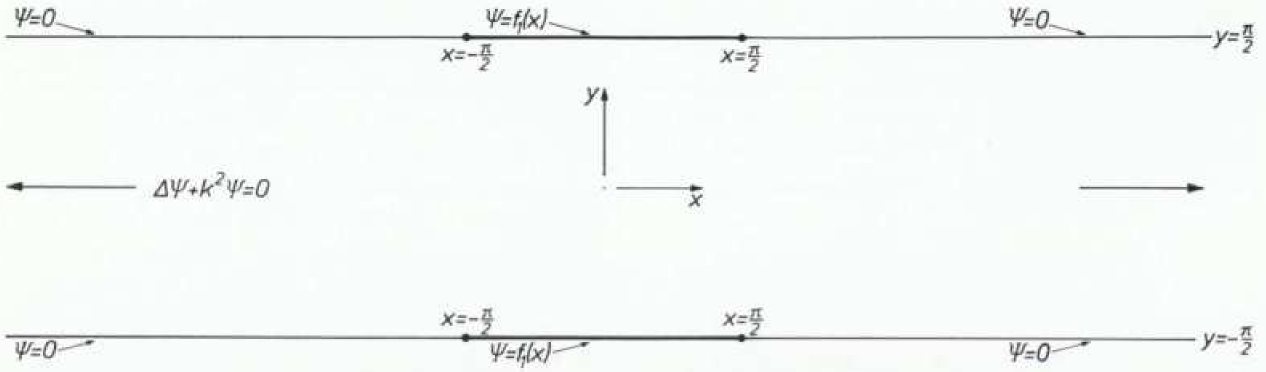


Fig. 3. Ter illustratie van het even randwaardeprobleem.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

$-\frac{1}{2}\pi < x < \frac{1}{2}\pi$. Stel dat $\Psi(x,y)$ een oplossing is van de golfvergelijking binnen het gebied dat wordt begrensd door de twee horizontale lijnen met $\Psi = 0$ op de rand, behalve voor $y = \pm\frac{1}{2}\pi$, $-\frac{1}{2}\pi < x < \frac{1}{2}\pi$, waar geldt: $\Psi = f_1(x)$. Laat verder $\Psi(x,y)$ even in x en even in y zijn, en laat Ψ ten slotte voldoen aan de stralingsconditie in $x = \pm\infty$. Deze functie $\Psi(x,y)$ bestaat en is eenduidig bepaald indien wordt voldaan aan zekere voorwaarden van gladheid, welke hier niet worden gespecificeerd. Zij wordt gegeven door:

$$\Psi(x,y) := \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} f_1(t)\,\mathrm{d}t \times$$

$$\times \frac{1}{\pi}\int_0^\infty \frac{\cosh\{y\sqrt{(s^2-k^2)}\}}{\cosh\{\frac{1}{2}\pi\sqrt{(s^2-k^2)}\}}\cos\{s(x-t)\}\,\mathrm{d}s. \quad (1)$$

Ten einde dit te bewijzen, merken we op dat het rechterlid (RL) van (1) een even functie van y is. Het is ook een even functie van x, omdat we in (1) de uitdrukking $\cos\{s(x-t)\} = \cos(sx)\cos(st) + \sin(sx)\sin(st)$ kunnen vervangen door $\cos(sx)\cos(st)$ zonder de integraal te wijzigen (merk op dat f_1 even is). Aldus heeft $\Psi(x,y)$ de juiste symmetrie. Wegens het theorema van Fourier gaat verder RL over in $f_1(x)$ voor $y = \pm\frac{1}{2}\pi$ indien geldt $-\frac{1}{2}\pi < x < \frac{1}{2}\pi$, en is RL buiten dit interval nul. Dientengevolge voldoet $\Psi(x,y)$ aan de juiste randvoorwaarden. Vervolgens is, binnen de strook van fig. 3, RL inderdaad een oplossing van de golfvergelijking omdat $\cosh\{y\sqrt{(s^2-k^2)}\}\cos\{s(x-t)\}$ dit is voor iedere waarde van s en t. Deze eigenschap blijft geldig na integratie over s en t, dit na vermenigvuldiging met gewichtsfactoren die onafhankelijk zijn van x en y. Om in te zien dat RL voldoet aan de stralingsconditie in het oneindige, is een wat omvangrijker analyse nodig. Laten wij hiertoe voor reële t de functie (even in t) beschouwen, die als volgt is gedefinieerd:

$$\chi(t) := \frac{1}{\pi}\int_0^\infty \frac{\cosh\{y\sqrt{(s^2-k^2)}\}}{\cosh\{\frac{1}{2}\pi\sqrt{(s^2-k^2)}\}}\cos(st)\,\mathrm{d}s.$$

Onafhankelijk van de keus van de vierkantswortel is de integrand analytisch en een even functie van s. Daarom geldt:

$$\chi(t) = \frac{1}{2\pi}\int_{-\infty}^\infty \frac{\cosh\{y\sqrt{(s^2-k^2)}\}}{\cosh\{\frac{1}{2}\pi\sqrt{(s^2-k^2}\}}\exp[is|t|]\,\mathrm{d}s.$$

De integrand heeft oneindig veel enkelvoudige polen in $s = \sqrt{(k^2-n^2)}$ (n oneven, > 0) in het bovenhalfvlak van de complexe variabele s, met residuen

$$\frac{2}{\pi}\,\frac{n}{\sqrt{(k^2-n^2)}}\sin(\tfrac{1}{2}n\pi)\cos(ny)\exp[i|t|\sqrt{(k^2-n^2)}].$$

Dus geldt:

$$\frac{1}{\pi}\int_0^\infty \frac{\cosh\{y\sqrt{(s^2-k^2)}\}}{\cosh\frac{1}{2}\{\pi\sqrt{(s^2-k^2)}\}}\cos\{s(x-t)\}\,\mathrm{d}s = \frac{2i}{\pi}\sum_{\substack{n\ \text{oneven}\\ >0}} \frac{n}{\sqrt{(k^2-n^2)}}\sin(\tfrac{1}{2}n\pi)\cos(ny)\exp[i|x-t|\sqrt{(k^2-n^2)}].$$

Het linkerlid hiervan is een oplossing van de golfvergelijking; het rechterlid is de bijbehorende ontwikkeling in modi, waarvan iedere term voldoet aan de stralingsconditie in het oneindige: de eerste paar representeren voortlopende golven die naar het oneindige stralen, en de overblijvende modi representeren exponentieel gedempte golven voor $|x| \to \infty$. Na vermenigvuldiging met $f_1(t)$ en integratie over $-\frac{1}{2}\pi \leqslant t \leqslant \frac{1}{2}\pi$ blijven deze eigenschappen geldig voor $|x| > \frac{1}{2}\pi$. Aldus voldoet RL van vergelijking (1) inderdaad aan de vereiste stralingsconditie.

Dit besluit ons bewijs dat de functie $\Psi(x,y)$, gedefinieerd door vergelijking (1), de (eenduidige) oplossing van ons randwaardeprobleem van Dirichlet is.

Zij nu $\phi_1(x,y)$ het totale veld van het even probleem en stel dat $f_1(x)$ en $g_1(x)$ de waarden van ϕ_1 zijn op de rand van gebied V (zie *fig. 4*); f_1 en g_1 zijn beide even functies van hun respectieve argumenten. Merk op dat de oorspronkelijke golf niet bijdraagt tot f_1, omdat deze voor $y = \pm\frac{1}{2}\pi$ identiek nul is.

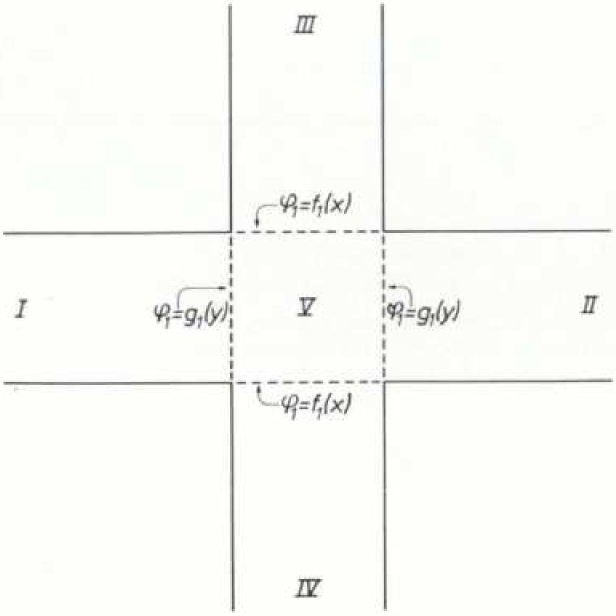


Fig. 4. Randwaarden voor het even probleem.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

In de vereniging van de gebieden *I*, *V* en *II* hebben we:

$$\phi_1(x,y) = \exp[\tfrac{1}{2}\pi i \sqrt{(k^2-1)}]\cos y \cos\{x\sqrt{(k^2-1)}\} + \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} f_1(t)\,\mathrm{d}t\, \frac{1}{\pi}\int_0^\infty \frac{\cosh\{y\sqrt{(s^2-k^2)}\}}{\cosh\{\frac{1}{2}\pi\sqrt{(s^2-k^2)}\}}\cos\{s(x-t)\}\,\mathrm{d}s,$$

wegens de representatie-stelling (1).

Anderzijds hebben we in de vereniging van de gebieden *III*, *V* en *IV*:

$$\phi_1(x,y) = \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} g_1(t)\,\mathrm{d}t\, \frac{1}{\pi}\int_0^\infty \frac{\cosh\{x\sqrt{(s^2-k^2)}\}}{\cosh\{\frac{1}{2}\pi\sqrt{(s^2-k^2)}\}}\cos\{s(y-t)\}\,\mathrm{d}s,$$

door toepassen van vergelijking (1) op de „verticale" armen; merk op dat x en y zijn verwisseld.

De gebieden waar deze twee vergelijkingen geldig zijn overdekken elkaar in gebied *V*, de rand hiervan inbegrepen. De eerste vergelijking toepassen op $x = \frac{1}{2}\pi$ geeft:

$$g_1(y) = \tfrac{1}{2}(\exp[\pi i\sqrt{(k^2-1)}]+1)\cos y + \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} f_1(t)\,\mathrm{d}t\, \frac{1}{\pi}\int_0^\infty \frac{\cosh\{y\sqrt{(s^2-k^2)}\}}{\cosh\{\frac{1}{2}\pi\sqrt{(s^2-k^2)}\}}\cos\{s(\tfrac{1}{2}\pi-t)\}\,\mathrm{d}s\,,$$

en toepassen van de tweede vergelijking op $y = \frac{1}{2}\pi$ geeft:

$$f_1(x) = \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} g_1(t)\,\mathrm{d}t\, \frac{1}{\pi}\int_0^\infty \frac{\cosh\{x\sqrt{(s^2-k^2)}\}}{\cosh\{\frac{1}{2}\pi\sqrt{(s^2-k^2)}\}}\cos\{s(\tfrac{1}{2}\pi-t)\}\,\mathrm{d}s\,.$$

Deze laatste twee vergelijkingen vormen een stelsel van gekoppelde lineaire integraalvergelijkingen voor de twee onbekende functies f_1 en g_1. Deze kunnen enigszins worden vereenvoudigd door gebruik te maken van de pariteitseigenschappen van f_1 en g_1: de factor $\cos\{s(\frac{1}{2}\pi-t)\}$ kan worden vervangen door $\cos(\frac{1}{2}\pi s)\cos(st)$. Definieer, om de notatie te vereenvoudigen, de kern voor het even probleem door

$$K(x,t) := \frac{1}{\pi}\int_0^\infty \frac{\cosh\{x\sqrt{(s^2-k^2)}\}}{\cosh\{\frac{1}{2}\pi\sqrt{(s^2-k^2)}\}}\cos(\tfrac{1}{2}\pi s)\cos(st)\,\mathrm{d}s\,. \qquad (2)$$

We krijgen dan de integraalvergelijkingen voor het even probleem in de volgende vorm:

$$\left.\begin{aligned} f_1(x) &= \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} K(x,t)\,g_1(t)\,\mathrm{d}t \qquad (-\tfrac{1}{2}\pi < x < \tfrac{1}{2}\pi),\\ g_1(y) &= \tfrac{1}{2}(\exp[\pi i\sqrt{(k^2-1)}]+1)\cos y + \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} K(y,t)\,f_1(t)\,\mathrm{d}t\,(-\tfrac{1}{2}\pi < y < \tfrac{1}{2}\pi).\end{aligned}\right\} \qquad (3)$$

We richten ons nu op het oneven probleem. Het totale veld zal worden aangeduid met $\phi_2(x,y)$, welke functie even is in y, maar oneven in x. De waarden van ϕ_2 op de rand van gebied *V* worden aangeduid met $f_2(x)$ en $\pm g_2(y)$, waarbij g_2 even is in y en f_2 oneven is in x (zie *fig. 5*).

In de vereniging van de gebieden *I*, *V* en *II* hebben we:

$$\phi_2(x,y) = i\exp[\tfrac{1}{2}\pi i\sqrt{(k^2-1)}]\cos y \sin\{x\sqrt{(k^2-1)}\} + \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} f_2(t)\,\mathrm{d}t\, \frac{1}{\pi}\int_0^\infty \frac{\cosh\{y\sqrt{(s^2-k^2)}\}}{\cosh\{\frac{1}{2}\pi\sqrt{(s^2-k^2)}\}}\cos\{s(x-t)\}\,\mathrm{d}s\,,$$

dit weer door toepassing van de representatie-stelling (1).

Anderzijds hebben we in de vereniging van de gebieden *III*, *V* en *IV*:

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

$$\phi_2(x,y) = \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} g_2(t)\,\mathrm{d}t\;\frac{1}{\pi}\int_0^\infty \frac{\sinh\{x\,\sqrt{(s^2-k^2)}\}}{\sinh\{\frac{1}{2}\pi\,\sqrt{(s^2-k^2)}\}}\cos\{s(y-t)\}\,\mathrm{d}s\,,$$

hetgeen is verkregen met behulp van een representatie-stelling die volkomen analoog is met die van vergelijking (1).

Deze twee representaties overdekken elkaar in gebied V, en door ze op de rand van gebied V toe te passen, verkrijgen we weer een stelsel van twee simultane integraalvergelijkingen. Het verschil met het even probleem is dat we nu twee kernen nodig hebben.

Deze twee kernen worden als volgt gedefinieerd:

$$L(x,t) := \frac{1}{\pi}\int_0^\infty \frac{\cosh\{x\,\sqrt{(s^2-k^2)}\}}{\cosh\{\frac{1}{2}\pi\,\sqrt{(s^2-k^2)}\}}\sin(\tfrac{1}{2}\pi s)\sin(st)\,\mathrm{d}s\,, \qquad (4)$$

$$M(x,t) := \frac{1}{\pi}\int_0^\infty \frac{\sinh\{x\,\sqrt{(s^2-k^2)}\}}{\sinh\{\frac{1}{2}\pi\,\sqrt{(s^2-k^2)}\}}\cos(\tfrac{1}{2}\pi s)\cos(st)\,\mathrm{d}s\,. \qquad (5)$$

Het resulterende stelsel van integraalvergelijkingen voor het oneven probleem is dan:

$$\left.\begin{aligned} f_2(x) &= \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} M(x,t)\,g_2(t)\,\mathrm{d}t \qquad (-\tfrac{1}{2}\pi < x < \tfrac{1}{2}\pi),\\ g_2(y) &= \tfrac{1}{2}\,(\exp[\pi i\,\sqrt{(k^2-1)}]-1)\cos y + \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} L(y,t)\,f_2(t)\,\mathrm{d}t \quad (-\tfrac{1}{2}\pi < y < \tfrac{1}{2}\pi).\end{aligned}\right\} \qquad (6)$$

Natuurlijk hebben we niet geprobeerd de stelsels (3) en (6) van gekoppelde lineaire vergelijkingen in gesloten vorm op te lossen, indien dit al mogelijk ware. Bovendien zal de technicus niet geïnteresseerd zijn in numerieke tabellen van de onbekende functies f en g, zodat een directe numerieke oplossing van (3) en (6) met behulp van de computer van weinig belang is. De technicus is wèl geïnteresseerd in een tabel van waarden van de reflectie-, transmissie- en deflectie-coëfficiënten, zoals die in het vorige hoofdstuk zijn gedefinieerd, omdat deze een duidelijke fysische betekenis hebben en experimenteel niet moeilijk te meten zijn. Gelukkig zijn deze coëfficiënten eenvoudige combinaties van die van de Fourier-ontwikkelingen van de functies f en g. Aldus kunnen we zowel (3) als (6) omzetten in een oneindig stelsel van gekoppelde inhomogene lineaire algebraïsche vergelijkingen met respectievelijk een oneindig groot aantal onbekenden A_n, B_n en C_n, D_n. Deze omzetting wordt besproken in het volgende hoofdstuk.

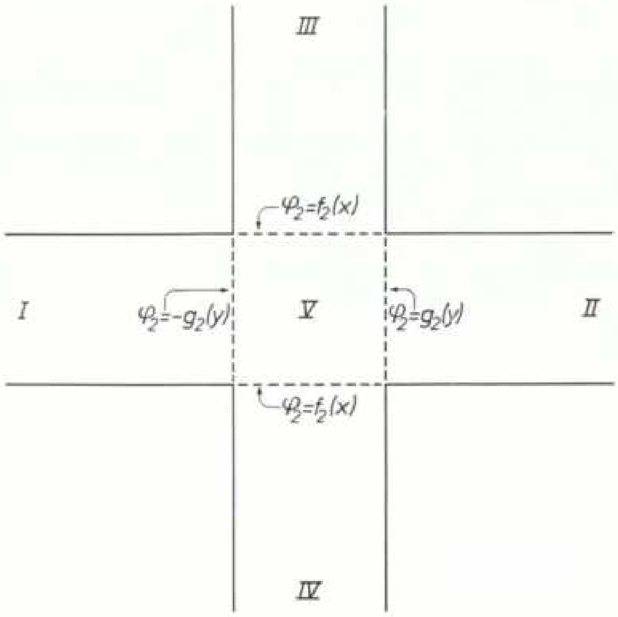


Fig. 5. Randwaarden voor het oneven probleem.

Fourier-ontwikkeling

Allereerst kunnen de drie kernen $K(x,t)$, $L(x,t)$ en $M(x,t)$ worden ontwikkeld in Fourier-reeksen naar x voor $|x| \leqslant \frac{1}{2}\pi$ en $|t| \leqslant \frac{1}{2}\pi$. Dit kan worden bereikt door middel van contourintegratie en residu-berekening. We zullen niet in bijzonderheden treden, maar alleen de uiteindelijke resultaten noemen:

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

$$K(x,t) = \frac{2i}{\pi} \sum_{n \text{ oneven} > 0} \frac{n}{\sqrt{(k^2-n^2)}} \exp[\tfrac{1}{2}\pi i \sqrt{(k^2-n^2)}] \sin(\tfrac{1}{2}n\pi) \cos(nx) \cos\{t\sqrt{(k^2-n^2)}\}, \tag{7}$$

$$L(x,t) = \frac{2}{\pi} \sum_{n \text{ oneven} > 0} \frac{n}{\sqrt{(k^2-n^2)}} \exp[\tfrac{1}{2}\pi i \sqrt{(k^2-n^2)}] \sin(\tfrac{1}{2}n\pi) \cos(nx) \sin\{t\sqrt{(k^2-n^2)}\}, \tag{8}$$

$$M(x,t) = \frac{2}{\pi i} \sum_{n \text{ even} > 0} \frac{n}{\sqrt{(k^2-n^2)}} \exp[\tfrac{1}{2}\pi i \sqrt{(k^2-n^2)}] \cos(\tfrac{1}{2}n\pi) \sin(nx) \cos\{t\sqrt{(k^2-n^2)}\}. \tag{9}$$

Ten tweede voeren we de onbekende Fourier-coëfficiënten van de functies f_1, g_1, f_2, g_2 als volgt in (vgl. de coëfficiënten in tabel I):

$$f_1(x) = \sum_{n \text{ oneven} > 0} A_n \sin(\tfrac{1}{2}n\pi) \cos(nx), \tag{10}$$

$$g_1(y) = \tfrac{1}{2}(\exp[\pi i \sqrt{(k^2-1)}] + 1) \cos y + \sum_{n \text{ oneven} > 0} B_n \sin(\tfrac{1}{2}n\pi) \cos(ny), \tag{11}$$

$$f_2(x) = -\sum_{n \text{ even} > 0} C_n \cos(\tfrac{1}{2}n\pi) \sin(nx), \tag{12}$$

$$g_2(y) = \tfrac{1}{2}(\exp[\pi i \sqrt{(k^2-1)}] - 1) \cos y + \sum_{n \text{ oneven} > 0} D_n \sin(\tfrac{1}{2}n\pi) \cos(ny). \tag{13}$$

Deze vergelijkingen zijn geldig voor waarden van $|x|$ of $|y|$ die niet groter zijn dan $\frac{1}{2}\pi$.

Bij wijze van voorbeeld zullen we aangeven hoe (3) kan worden omgezet in algebraïsche vergelijkingen voor de coëfficiënten A_n en B_n van de Fourier-reeksen van f_1 en g_1.

Door gebruik te maken van (7) en (10) hebben we:

$$\int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} K(y,t) f_1(t)\, dt =$$

$$= \frac{2i}{\pi} \sum_{m \text{ oneven} > 0} A_m \sin(\tfrac{1}{2}m\pi) \sum_{n \text{ oneven} > 0} \frac{n}{\sqrt{(k^2-n^2)}} \exp[\tfrac{1}{2}\pi i \sqrt{(k^2-1)}] \sin(\tfrac{1}{2}n\pi) \cos(ny) \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \cos(mt) \cos\{t\sqrt{(k^2-n^2)}\}\, dt.$$

De overblijvende integraal is elementair:

$$\int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \cos(mt) \cos\{t\sqrt{(k^2-n^2)}\}\, dt = \frac{2m \sin(\tfrac{1}{2}m\pi) \cos\{\tfrac{1}{2}\pi \sqrt{(k^2-n^2)}\}}{m^2+n^2-k^2},$$

waarbij m oneven en positief is.

Dientengevolge gaat de coëfficiënt van $\sin(\frac{1}{2}n\pi)\cos(ny)$ voor vaste n over in

$$\frac{4i}{\pi} \frac{n}{\sqrt{(k^2-n^2)}} \exp[\tfrac{1}{2}\pi i \sqrt{(k^2-n^2)}] \cos\{\tfrac{1}{2}\pi \sqrt{(k^2-n^2)}\} \sum_{m \text{ oneven} > 0} \frac{m}{m^2+n^2-k^2} A_m.$$

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Tabel II. De twee stelsels lineaire vergelijkingen.

$$B_n = \frac{2i}{\pi}\frac{n}{\sqrt{(k^2-n^2)}}(1+\exp[\pi i\sqrt{(k^2-n^2)}])\sum_{m\text{ oneven}>0}\frac{m}{m^2+n^2-k^2}A_m \qquad (n\text{ oneven}>0) \qquad (14)$$

$$A_n = \frac{2i}{\pi}\frac{n}{\sqrt{(k^2-n^2)}}(1+\exp[\pi i\sqrt{(k^2-n^2)}])\left[\frac{\exp[\pi i\sqrt{(k^2-1)}]+1}{2(n^2+1-k^2)}+\sum_{m\text{ oneven}>0}\frac{m}{m^2+n^2-k^2}B_m\right] \qquad (n\text{ oneven}>0) \qquad (15)$$

$$D_n = \frac{2i}{\pi}\frac{n}{\sqrt{(k^2-n^2)}}(1-\exp[\pi i\sqrt{(k^2-n^2)}])\sum_{m\text{ even}>0}\frac{m}{m^2+n^2-k^2}C_m \qquad (n\text{ oneven}>0) \qquad (16)$$

$$C_n = \frac{2i}{\pi}\frac{n}{\sqrt{(k^2-n^2)}}(1+\exp[\pi i\sqrt{(k^2-n^2)}])\left[\frac{\exp[\pi i\sqrt{(k^2-1)}]-1}{2(n^2+1-k^2)}+\sum_{m\text{ oneven}>0}\frac{m}{m^2+n^2-k^2}D_m\right] \qquad (n\text{ even}>0) \qquad (17)$$

Wegens (11) en de tweede vergelijking van (3) moet deze coëfficiënt B_n zijn. Aldus krijgen we vergelijking (14) van *tabel II*.

Evenzo kunnen we door substitutie van (11) in de eerste vergelijking van (3) en gebruik makend van (7), A_n uitdrukken als lineaire combinatie van de coëfficiënten B, afgezien van de extra-term die verband houdt met het invallende veld. Als resultaat wordt gevonden vergelijking (15) van tabel II.

De omzetting van het stelsel (6) van integraalvergelijkingen gaat volkomen analoog. Het resultaat bestaat uit de vergelijkingen (16) en (17) van tabel II.

Men ziet dat de twee stelsels (14, 15) en (16, 17) grote overeenkomst vertonen. In feite zijn zij identiek wat betreft de numerieke aanpak en de computerprogrammering. Om dit in te zien voeren we de pariteitsindex j in, met $j = 0$ voor het even probleem en $j = 1$ voor het oneven probleem, en we stellen voor $n = 1, 2, 3, \ldots$:

$$A_{2n-1} = X(1,n), \qquad B_{2n-1} = Y(1,n),$$
$$C_{2n} = X(2,n), \qquad D_{2n-1} = Y(2,n).$$

Zoals gemakkelijk is te verifiëren, gaan de twee stelsels als volgt in één stelsel over:

$$Y(j+1,n) = -\frac{2(2n-1)}{\pi i\sqrt{\{k^2-(2n-1)^2\}}}\left[1+(-1)^j\exp[\pi i\sqrt{\{k^2-(2n-1)^2\}}]\right]\times$$
$$\times\sum_{m=1}^{\infty}\frac{2m-1+j}{(2m-1+j)^2+(2n-1)^2-k^2}X(j+1,m), \qquad (18)$$

$$X(j+1,n) = -\frac{2(2n-1+j)}{\pi i\sqrt{\{k^2-(2n-1+j)^2\}}}\left[1+\exp[\pi i\sqrt{\{k^2-(2n-1+j)^2\}}]\right]\times$$
$$\times\left[\frac{\exp[\pi i\sqrt{(k^2-1)}]+(-1)^j}{2\{(2n-1+j)^2+1-k^2\}}+\sum_{m=1}^{\infty}\frac{2m-1}{(2m-1)^2+(2n-1+j)^2-k^2}Y(j+1,m)\right]. \qquad (19)$$

Dit oneindige stelsel van simultane lineaire vergelijkingen voor de onbekenden X en Y kan gemakkelijk worden omgezet in één enkel stelsel voor X alleen, en wel door elimineren van Y. Voor numerieke berekeningen wordt dit stelsel afgekapt, b.v. bij $n = 30$, en het dan ontstane eindige stelsel van de orde 30 wordt opgelost met behulp van een van de bekende technieken. Zodra we de eerste 30 componenten van X kennen, berekenen we de eerste 30 componenten van Y met behulp van (18). Natuurlijk zijn de dan verkregen waarden benaderingen van de echte waarden.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Inderdaad zal de benadering zeer slecht zijn voor de laatste paar componenten, omdat deze sterk worden beïnvloed door het afkappen. Daarom zullen we de berekening beperken tot de eerste 15 componenten van X en Y.

Numerieke aanpak en programmering

Het is onze bedoeling een tabel te maken van modulus en argument van de coëfficiënten R_n, T_n en S_n $n = 1(1)30$ en $k = 1(0.1)13$ (k niet geheel). Als de gegevens voor één vaste k op één bladzijde worden afgedrukt, omvat deze tabel 108 bladzijden. We zullen ook de totale energie die wordt teruggekaatst, doorgelaten of afgebogen in de armen van de kruiskoppeling berekenen en afdrukken.

In het programma (in FORTRAN CDC-3600; zie het aanhangsel) gebruiken we NCP voor het bijhouden van het aantal coëfficiënten dat afgedrukt moet worden (NCP is ten hoogste 15) en NEQ voor de orde van het eindige stelsel van lineaire vergelijkingen na afkappen van het oneindige stelsel. Verder stelt PI het getal π voor (in 10 cijfers), CI representeert de imaginaire eenheid i, KU is het aantal gehelen in k, KT het aantal tienden, KK representeert k^2, en k wordt zelf gerepresenteerd door K. In de benodigde waarden van k wordt voorzien door een dubbele DO-lus (DO 1).

Uit de vergelijkingen (18) en (19) blijkt dat voor $n = 1(1)60$ de waarden van $\pi i \sqrt{(k^2 - n^2)}$ en $\exp [\pi i \sqrt{(k^2 - n^2)}]$ nodig zijn. Het is dus verstandig deze eens en voor al te berekenen en op te slaan. We representeren de hiermee corresponderende velden (Engels: arrays) met behulp van respectievelijk RT (afkorting van „root") en ER (exponentiële functie van „root") (DO 2). Merk op dat AR het argument $(k^2 - n^2)$ van de wortel representeert. Bij IF wordt er een controle op de waarde van AR uitgevoerd. Als $k^2 - n^2 > 0$ is, springt het programma naar de opdracht met label 21. Dit wil zeggen: voor $k > n$ krijgen we $\mathrm{RT}(n) = \pi i(k^2 - n^2)^{\frac{1}{2}}$. Anderzijds, als $k < n$ is (merk op dat $k = n$ niet mogelijk is omdat k niet geheel is), wordt de opdracht RT(N) = —PI★SQRTF(—AR) uitgevoerd en het programma springt naar de opdracht met label 22, aldus opdracht 21 overslaand. Voor $k < n$ geldt dus: $\mathrm{RT}(n) = -\pi(k^2 - n^2)^{\frac{1}{2}}$. Deze berekening van $\pi i \sqrt{(k^2 - n^2)}$ is exact in overeenstemming met de hierboven gegeven definitie van de vierkantswortel. In beide gevallen wordt het resultaat als complex getal opgeslagen. Opdracht 22 dient ook om de hiermee corresponderende juiste waarde van de exponentiële uitdrukking te berekenen en op te slaan, hierbij gebruik makend van de bibliotheek-subroutine CEXP.

Vervolgens zouden we eraan willen herinneren dat er twee problemen zijn die moeten worden opgelost: het even probleem en het oneven probleem. De computer maakt hiertoe gebruik van de pariteitsindex J, die j representeert. Dit leidt tot een zeer eenvoudige lus (DO 3).

In dit stadium van het programmeren kijken we opnieuw naar de vergelijkingen (18) en (19). Zij kunnen als volgt worden geschreven:

$$Y(j+1,n) = f_1(n) \sum_{m=1}^{\infty} \frac{2m-1+j}{(2m-1+j)^2 + (2n-1)^2 - k^2} X(j+1,m),$$

$$X(j+1,n) = -f(n) + f_2(n) \sum_{m=1}^{\infty} \frac{2m-1}{(2n-1+j)^2 + (2m-1)^2 - k^2} Y(j+1,m),$$

waarin

$$f_1(n) = -\frac{2(2n-1)}{\mathrm{RT}(2n-1)} \{1 + (-1)^j \,\mathrm{ER}(2n-1)\},$$

$$f_2(n) = -\frac{2(2n-1+j)}{\mathrm{RT}(2n-1+j)} \{1 + \mathrm{ER}(2n-1+j)\},$$

$$f(n) = -f_2(n) \frac{\mathrm{ER}(1) + (-1)^j}{2\{(2n-1+j)^2 + 1 - k^2\}}.$$

We hebben de corresponderende waarden nodig voor $n = 1(1)30$; we berekenen deze daarom en slaan ze op (DO 4). De waarden voor $j = 0$, die het eerst worden berekend en opgeslagen, worden dan eerst gebruikt en later overschreven door de waarden voor $j = 1$.

We komen nu toe aan het elimineren van Y en de berekening van de matrixelementen en de rechterleden van de hierdoor ontstane vergelijkingen in X. In feite blijkt het rechterlid de functie $f(n)$ te zijn. Als de vergelijkingen worden geschreven in de vorm

$$\sum_{m=1}^{\infty} \mathrm{MAT}(n,m)\, X(m) = f(n),$$

is het niet moeilijk te laten zien dat geldt:

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

$$\text{MAT}(n,m) = -\delta_{n,m} + (2m-1+j)f_2(n) \sum_{s=1}^{\infty} \frac{(2s-1)f_1(s)}{\{(2s-1)^2+(2n-1+j)^2-k^2\}\{(2s-1)^2+(2m-1+j)^2-k^2\}}.$$

De oneindige som wordt natuurlijk afgebroken bij s = NEQ. Merk ook op dat MAT niet symmetrisch is. Het hiermee corresponderende programmagedeelte omvat een drievoudige DO-lus. Eerst is er een (DO 5) voor $n = 1(1)30$, dan een (DO 6) voor $m = 1(1)30$, en tenslotte een (DO 7) voor de berekening van de som over $s = 1(1)30$. Hierbij is 30 de maximaal toegestane waarde van NEQ. De hierbij voorkomende lusvariabelen zijn N, M en I.

Thans zijn we klaar om de lineaire vergelijkingen op te lossen, omdat de matrix en het rechterlid in het geheugen beschikbaar zijn. Dit gebeurt door middel van een onafhankelijk onderprogramma, dat in dit stadium door het hoofdprogramma wordt aangeroepen.

Na terugkeer uit het onderprogramma CLINEQ (afkorting van „complex linear equations"), is de oplossing in het veld XX beschikbaar. Het hoofdprogramma moet dan deze oplossing opslaan in het veld X(NC,N), hetgeen bereikt wordt door middel van een enkelvoudige lus (DO 8).

We hebben dan vergelijking (18) nodig, ten einde de corresponderende elementen van het Y-veld te berekenen en op te slaan. Dit wordt gedaan met behulp van een dubbele lus (DO 9 en DO 10).

We zijn nu met het programma zover dat we de NC-lus kunnen sluiten. Als de computer deze lus (DO 3) heeft volbracht, hebben we X(1,N), Y(1,N), X(2,N) en Y(2,N) numeriek beschikbaar voor verder gebruik, waarbij N = 1(1)30. We nemen aan dat de eerste 15 componenten werkelijk significant zijn en laten de andere 15 weg. Vervolgens berekenen we de reflectiecoëfficiënten R_{2n-1}, de transmissiecoëfficiënten T_{2n-1} en de deflectiecoëfficiënten $S_{2n-1} = A_{2n-1}$ en $S_{2n} = -C_{2n}$ voor $n = 1(1)15$ en slaan deze op. De hiermee corresponderende velden worden aangeduid met REF, TRM en DEF (DO 11). De opdracht die volgt na deze lus is nodig om de invallende golf in gebied *II* op te nemen in de doorgelaten dominante modus.

We zijn geïnteresseerd in de absolute waarden en fasen van de verschillende coëfficiënten (DO 12 en DO 13)

Bij een numeriek probleem als het onderhavige is de graad van nauwkeurigheid van de resultaten die men met het programma verkrijgt moeilijk te schatten. In dit geval kan de energiebalans wat helpen. De invallende energie, die wordt aangevoerd door de exciterende dominante modus, moet theoretisch gelijk zijn aan de som van de teruggekaatste, doorgelaten en zijwaarts afgebogen energieën. Deze worden alleen bepaald door de ongedempte modi. Nu is de energiestroom in de x-richting evenredig met het imaginaire gedeelte van $\bar{\Phi}\,\partial\Phi/\partial x$, geïntegreerd over de doorsnede van de golfgeleider (met $\bar{\Phi}$ wordt de toegevoegd complexe van Φ bedoeld). Laten we de totale energie die per seconde op de kruiskoppeling valt gelijk aan één stellen. Dan geldt bij voorbeeld in gebied *I* voor de gereflecteerde energie:

$$\frac{1}{\sqrt{(k^2-1)}} \sum_{0<n \text{ oneven} < k} |T_n|^2 \sqrt{(k^2-n^2)},$$

hetgeen kan worden herschreven als

$$\frac{1}{\text{RT}(1)} \sum_{0<n \text{ oneven} < k} |T_n|^2\, \text{RT}(n).$$

Verder kunnen we, daar de gedempte golven niet tot de som bijdragen, de sommatie uitstrekken over alle oneven waarden van n tot NEQ. Dergelijke uitdrukkingen gelden ook voor de energie die doorgelaten wordt in gebied *II* en voor de energie die zijwaarts afgebogen wordt in de gebieden *III* en *IV*. Aangezien deze energieën van fysisch belang zijn, berekenen we ze, om ze later te kunnen afdrukken. De variabelen zijn EREF, ETRM en EDEF. We berekenen ook SUMS, een grootheid die theoretisch gelijk aan één zou moeten zijn. De afwijking van één geeft een indruk van de nauwkeurigheid van het numerieke proces. Het is geen controle op het effect van het afkappen, zoals Boersma heeft laten zien: de energiebalans geldt voor iedere orde N na afkappen van het oneindige stelsel van lineaire vergelijkingen.

Nu alle van belang zijnde grootheden zijn berekend en opgeslagen, kunnen we één bladzijde met gegevens laten afdrukken (DO 17).

De dubbele lus van de variabelen KU en KT, die de waarden van k bijhouden, kan nu worden gesloten.

Het programma eindigt met de vereiste FORMAT-specificaties.

Het onderprogramma voor het oplossen van de lineaire vergelijkingen was ons welwillend ter beschikking gesteld door M. J. Geleijns van Philips' Rekencentrum.

Bespreking van de numerieke resultaten

In *tabel III* hebben we de numerieke gegevens bijeengebracht die afgedrukt zijn op de eerste regel van ieder van de 108 geproduceerde pagina's. Ter toelichting: r = 100 EREF is de totale hoeveelheid energie die wordt teruggekaatst in gebied *I* als percentage van de invallende energie, t = 100 ETRM is de hiermee correspon-

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

derende hoeveelheid in gebied *II* doorgelaten energie, en $s = 100$ EDEF is de energie die zijwaarts wordt verstrooid in ieder van de twee gebieden *III* en *IV*. Afgezien van afrondingsfouten geldt: $r + t + 2s = 100$. En passant merken we op dat SUMS = 1.000 000 is voor ieder van de 108 waarden van k. Een tweede uitvoering van het programma met andere waarden van k (in de buurt van gehele getallen; zie hieronder) heeft geleerd dat SUMS steeds tot in 10 decimalen gelijk is aan één.

Tabel III. De gereflecteerde, de doorgaande en de afgebogen energie als percentage van de invallende energie.

k	r	t	s	k	r	t	s	k	r	t	s
1.1	37.93	46.60	7.73	5.1	1.30	93.74	2.48	9.1	0.59	96.86	1.27
1.2	16.00	61.69	11.16	5.2	0.93	94.44	2.31	9.2	0.41	97.32	1.14
1.3	7.28	67.02	12.85	5.3	0.76	94.97	2.14	9.3	0.33	97.65	1.01
1.4	3.86	68.90	13.62	5.4	0.68	95.41	1.95	9.4	0.30	97.91	0.90
1.5	2.77	69.68	13.78	5.5	0.67	95.79	1.77	9.5	0.30	98.12	0.79
1.6	2.78	70.35	13.44	5.6	0.71	96.08	1.61	9.6	0.32	98.28	0.70
1.7	3.31	71.38	12.66	5.7	0.80	96.28	1.46	9.7	0.38	98.37	0.63
1.8	4.18	72.92	11.45	5.8	0.96	96.39	1.33	9.8	0.47	98.40	0.56
1.9	5.64	74.70	9.83	5.9	1.26	96.32	1.21	9.9	0.65	98.31	0.52
2.1	6.51	54.24	19.62	6.1	1.08	93.71	2.61	10.1	0.49	97.15	1.18
2.2	4.97	55.63	19.70	6.2	0.84	94.16	2.50	10.2	0.37	97.42	1.10
2.3	4.15	58.84	18.50	6.3	0.75	94.70	2.28	10.3	0.33	97.69	0.99
2.4	3.70	62.57	16.87	6.4	0.73	95.20	2.04	10.4	0.32	97.94	0.87
2.5	3.52	66.68	14.90	6.5	0.74	95.62	1.82	10.5	0.34	98.13	0.76
2.6	3.56	71.56	12.44	6.6	0.79	95.96	1.62	10.6	0.38	98.27	0.68
2.7	3.81	77.49	9.35	6.7	0.89	96.24	1.43	10.7	0.44	98.34	0.61
2.8	4.35	83.63	6.01	6.8	1.07	96.44	1.24	10.8	0.53	98.35	0.56
2.9	6.12	85.87	4.01	6.9	1.42	96.38	1.10	10.9	0.68	98.24	0.54
3.1	3.94	76.78	9.64	7.1	0.88	94.95	2.09	11.1	0.44	97.65	0.96
3.2	2.33	82.66	7.51	7.2	0.61	95.45	1.97	11.2	0.32	97.84	0.92
3.3	1.62	86.32	6.03	7.3	0.49	95.92	1.79	11.3	0.27	98.03	0.85
3.4	1.24	89.01	4.87	7.4	0.43	96.38	1.59	11.4	0.24	98.20	0.78
3.5	1.03	91.04	3.96	7.5	0.40	96.83	1.38	11.5	0.24	98.37	0.70
3.6	0.96	92.52	3.26	7.6	0.40	97.25	1.18	11.6	0.24	98.52	0.62
3.7	1.04	93.47	2.74	7.7	0.43	97.60	0.99	11.7	0.26	98.66	0.54
3.8	1.35	93.80	2.42	7.8	0.51	97.82	0.83	11.8	0.31	98.76	0.46
3.9	2.24	92.99	2.38	7.9	0.72	97.80	0.74	11.9	0.41	98.79	0.40
4.1	2.02	84.46	6.76	8.1	0.73	95.13	2.07	12.1	0.40	97.63	0.99
4.2	1.33	87.65	5.51	8.2	0.54	95.62	1.92	12.2	0.30	97.78	0.96
4.3	1.04	90.35	4.31	8.3	0.45	96.23	1.66	12.3	0.26	97.99	0.87
4.4	0.91	92.43	3.33	8.4	0.40	96.85	1.37	12.4	0.24	98.22	0.77
4.5	0.91	93.92	2.58	8.5	0.39	97.40	1.10	12.5	0.24	98.45	0.66
4.6	1.07	94.85	2.04	8.6	0.41	97.83	0.88	12.6	0.25	98.66	0.54
4.7	1.51	95.05	1.72	8.7	0.49	98.08	0.72	12.7	0.28	98.83	0.45
4.8	2.57	93.75	1.84	8.8	0.68	98.04	0.64	12.8	0.34	98.91	0.37
4.9	4.06	90.31	2.81	8.9	1.14	97.39	0.74	12.9	0.50	98.79	0.36

Bestudering van tabel III toont het voorkomen van resonantieverschijnselen als k zich dichtbij een geheel getal bevindt. Zoals uit de vergelijkingen (14) en (15) blijkt, kunnen A_{2N-1} en B_{2N-1} in absolute waarde groot zijn als k in de buurt van het oneven getal $2N - 1$ ligt, omdat de factor voor het sommatieteken voor $k = 2N - 1$ oneindig groot wordt. In tegenstelling hiermee is D_{2N-1} niet noodzakelijkerwijs groot, zoals uit vergelijking (16) volgt. Dientengevolge kunnen R_{2N-1}, T_{2N-1} en S_{2N-1} onbegrensd aangroeien als $k \to 2N - 1$. Evenzo wordt C_{2N} in absolute waarde groot als k dicht bij het even getal $2N$ ligt, en S_{2N} kan onbegrensd aangroeien als $k \to 2N$. Aangezien echter de absolute waarde van het kwadraat van deze grootheden, vermenigvuldigd met $\sqrt{(k^2 - n^2)}$, eindig moet zijn (voor respectievelijk $n = 2N - 1$ en $n = 2N$) omdat hierdoor energie wordt gerepresenteerd, gaan de grootheden niet sterker naar oneindig dan $(k - n)^{-1/4}$, als $k \to n$.

De kritische index, $2N - 1$ of $2N$, verdeelt het gebied van k-waarden in twee delen: is k kleiner dan de kritische index (d.w.z. beneden de afsnijfrequentie), dan wordt de desbetreffende modus gedempt, is k groter dan de kritische index (d.w.z. boven de afsnijfrequentie), dan plant de modus zich voort. We moeten daarom uitkijken naar een eventueel oneindig grote amplitude van de n-de modus als n gelijk is aan k bij de afsnijfrequentie. Als k een geheel getal is, kan het

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

oneindig worden zich ook voordoen door andere oorzaken dan de afsnijfrequentie, ofschoon het resonantie-effect, waarbij de amplitude van de modus hoge toppen of diepe minima als functie van k vertoont, misschien minder geprononceerd is.

Er dient nog te worden opgemerkt dat de noemers $m^2 + n^2 - k^2$ in de vergelijkingen (14) t/m (17) volkomen onschuldig zijn. Zij duiden een soort resonantie bij $k = 2$, 3 en 4. Anderzijds lijkt de transmissie iets te vertonen dat op een buigpunt lijkt bij even, gehele waarden van k, en er is een scherp minimum bij $k = 3$. Hetzelfde geldt voor de deflectie. De pieken bij $k = 3$ zijn, zoals men ziet, bijzonder scherp; de beschikbare gegevens zijn niet voldoende om de extreme waarden te schatten met een nauwkeurigheid beter dan vijf procent.

Tabel IV. De gereflecteerde, de doorgaande en de afgebogen energie als percentage van de invallende energie.

k	r	t	s	k	r	t	s	k	r	t	s
1.93	6.37	75.08	9.28	5.93	1.41	96.22	1.18	9.93	0.73	98.24	0.51
1.95	7.04	75.17	8.89	5.95	1.55	96.11	1.17	9.95	0.81	98.17	0.51
1.97	8.03	74.98	8.50	5.97	1.75	95.94	1.16	9.97	0.91	98.07	0.51
1.99	9.84	73.98	8.09	5.99	2.10	95.58	1.16	9.99	1.09	97.88	0.52
2.01	10.59	61.94	13.74	6.01	1.95	94.02	2.02	10.01	0.96	97.16	0.94
2.03	8.95	57.71	16.67	6.03	1.56	93.72	2.36	10.03	0.75	97.06	1.09
2.05	7.98	55.81	18.10	6.05	1.36	93.64	2.50	10.05	0.64	97.05	1.15
2.07	7.29	54.83	18.94	6.07	1.22	93.64	2.57	10.07	0.57	97.08	1.18
2.93	7.64	83.43	4.46	6.93	1.60	96.19	1.10	10.93	0.75	98.16	0.54
2.95	9.51	79.26	5.62	6.95	1.77	95.97	1.13	10.95	0.81	98.08	0.55
2.97	13.02	70.12	8.43	6.97	2.00	95.58	1.21	10.97	0.89	97.96	0.57
2.99	20.69	46.98	16.16	6.99	2.36	94.85	1.39	10.99	1.02	97.75	0.61
3.01	12.07	56.14	15.89	7.01	1.86	94.17	1.98	11.01	0.84	97.45	0.86
3.03	7.86	66.28	12.93	7.03	1.41	94.46	2.07	11.03	0.66	97.50	0.92
3.05	6.05	70.85	11.55	7.05	1.19	94.63	2.09	11.05	0.56	97.54	0.95
3.07	4.97	73.76	10.63	7.07	1.04	94.77	2.10	11.07	0.50	97.59	0.96
3.93	2.79	92.26	2.48	7.93	0.85	97.70	0.73	11.93	0.47	98.76	0.39
3.95	3.31	91.50	2.60	7.95	0.96	97.58	0.73	11.95	0.52	98.71	0.38
3.97	4.06	90.35	2.79	7.97	1.14	97.38	0.74	11.97	0.60	98.64	0.38
3.99	5.39	88.27	3.17	7.99	1.44	96.99	0.79	11.99	0.73	98.49	0.39
4.01	4.76	82.66	6.29	8.01	1.39	95.44	1.59	12.01	0.71	97.82	0.74
4.03	3.48	82.54	6.99	8.03	1.10	95.15	1.88	12.03	0.57	97.68	0.87
4.05	2.85	82.94	7.10	8.05	0.94	95.07	1.99	12.05	0.50	97.63	0.93
4.07	2.44	83.51	7.03	8.07	0.84	95.07	2.05	12.07	0.45	97.62	0.97
4.93	4.10	89.71	3.10	8.93	1.37	96.96	0.84				
4.95	3.92	89.70	3.19	8.95	1.54	96.59	0.93				
4.97	3.60	90.00	3.20	8.97	1.73	96.16	1.06				
4.99	3.15	90.61	3.12	8.99	1.89	95.68	1.22				
5.01	2.29	92.29	2.71	9.01	1.28	95.99	1.37				
5.03	1.92	92.86	2.61	9.03	0.97	96.32	1.36				
5.05	1.68	93.20	2.56	9.05	0.80	96.52	1.34				
5.07	1.50	93.45	2.53	9.07	0.70	96.68	1.31				

aan met gebied V, de eigenlijke koppeling. Deze resonantie wordt, als de noemer nul is (b.v. $k^2 = 2$, $n = m = 1$), gecompenseerd door de factor voor het sommatieteken die óók naar nul gaat.

Voor nauwkeurige interpolatie zijn de gegevens die tot nu toe zijn verkregen onvoldoende als k in de buurt van een geheel getal ligt. Daarom werd het programma een tweede keer uitgevoerd ten einde 88 pagina's te leveren voor k = 1.93, 1.95, 1.97, 1.99, 2.01, 2.03, 2.05, 2.07, waarden in de buurt van $k = 2$, enz. De hiermee corresponderende energieën zijn samengebracht in *tabel IV*.

Voor waarden van k tot 4 zijn deze resultaten ook weergegeven in *fig. 6*. De reflectie toont scherpe pieken

In *fig. 7* worden van boven naar beneden krommen getoond voor AREF (1), ADEF (1) en ATRM (1). De eerste kromme moet naar oneindig gaan als k van boven af tot 1 nadert. De pieken bij k = 2, 3, 4, . . . schijnen een eindige hoogte te hebben. De krommen voor de deflectie en de transmissie tonen weer een verschillend gedrag al naar k een even of een oneven geheel getal is.

Er is geen aanwijzing dat ATRM (1) naar oneindig gaat als k naar 1 gaat.

We hebben niet gepoogd de volledige inhoud van de numerieke tabellen in grafieken weer te geven. Zowel de schrijver van dit artikel als dr. Butterweck zijn in het bezit van een exemplaar van de numerieke tabel van

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

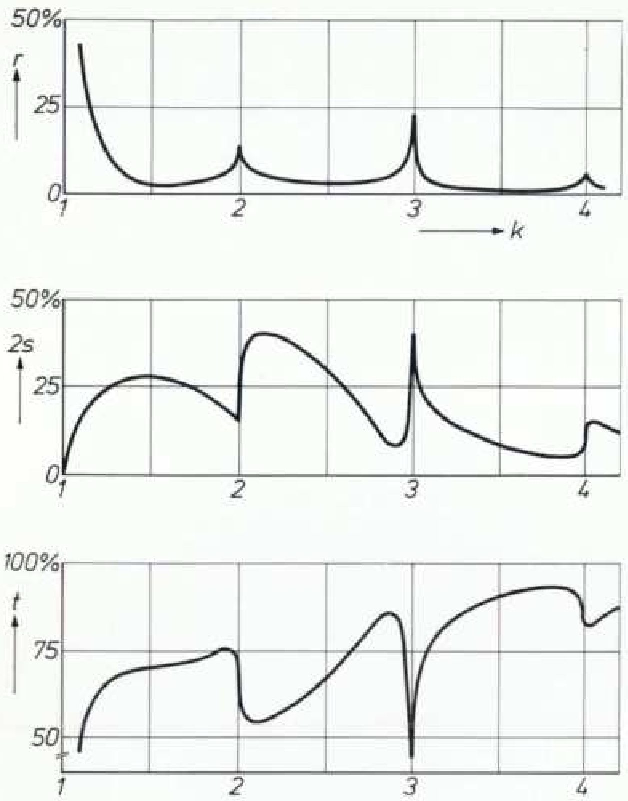


Fig. 6. De gereflecteerde (r), doorgaande (t) en afgebogen ($2s$) fracties van de invallende energie als functie van het golfgetal k.

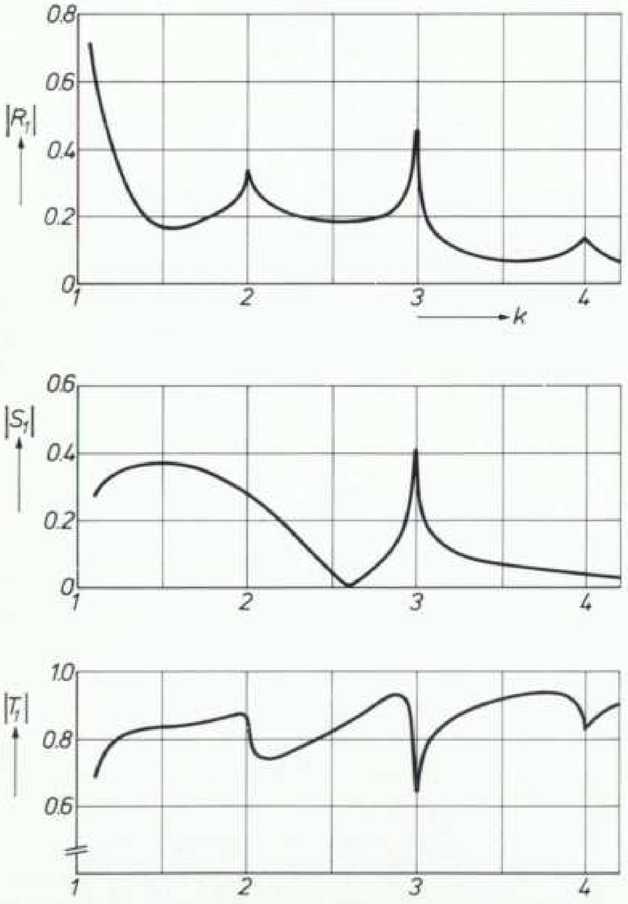


Fig. 7. Als fig. 6, maar nu voor de amplitude van de dominante modi.

196 pagina's, die, indien nodig, kunnen worden geraadpleegd.

Het programma vergde op de computer CDC-3600/3200 een rekentijd van minder dan 2 uur.

Aanhangsel: Computer-programma (FORTRAN CDC-3600)

```
    PRØGRAM CRØSS
    DIMENSIØN RT(60), ER(60), F1(30), F2(30), F(30),
              MAT(30, 30), X(2,30), Y(2,30), REF(15),
              TRM(15), DEF(30), XX(30), AREF(15),
              ATRM(15), ADEF(30), PREF(15), PTRM
              (15), PDEF(30)
    TYPE CØMPLEX CI, RT, ER, F1, F2, F, SUM, MAT,
              X, Y, REF, TRM, DEF, XX
    TYPE REAL K, KK

    NCP = 15
    NEQ = 2 ★NCP
    PI  = 3.141592654
    CI  = (0.,1.)
    DØ 1  KU = 1,12
    DØ 1  KT = 1,9
    K   = KU + 0.1 ★KT
    KK  = K ★K

    M = 2★NEQ
    DØ 2 N = 1,M
    AR = KK − N★N
    IF(AR.GT.0) GØTØ 21
    RT(N) = −PI ★SQRTF(−AR)
    GØTØ 22
 21 RT(N) = PI ★SQRTF(AR) ★CI
 22 ER(N) = CEXP(RT(N))
  2 CØNTINUE

    DØ 3 NC = 1,2
    J = NC − 1

    DØ 4 N = 1,NEQ
    NØ = 2★N − 1
    NE = NØ + J
    F1(N) = −2★NØ / RT(NØ) ★(1. + (−1) ★★J ★ER(NØ))
    F2(N) = −2★NE / RT(NE) ★(1. + ER(NE))
    F(N)  = −F2(N) ★(ER(1) + (−1) ★★J) / (2. ★(NE ★NE
           +1.−KK))
  4 CØNTINUE

    DØ 5 N = 1,NEQ
    NE = 2★N − 1 + J
    EN = NE★NE − KK
    DØ 6 M = 1,NEQ
    ME = 2★M − 1 + J
    EM = ME★ME − KK
    SUM = 0
    DØ 7 I = 1,NEQ
    L = 2★I − 1
    EL = L★L
    SUM = SUM + L★F1(I)/((EL + EN) ★(EL + EM))
  7 CØNTINUE
    MAT(N,M) = F2(N) ★ME ★SUM
  6 CØNTINUE
    MAT(N,N) = MAT(N,N) − 1.
  5 CØNTINUE
```

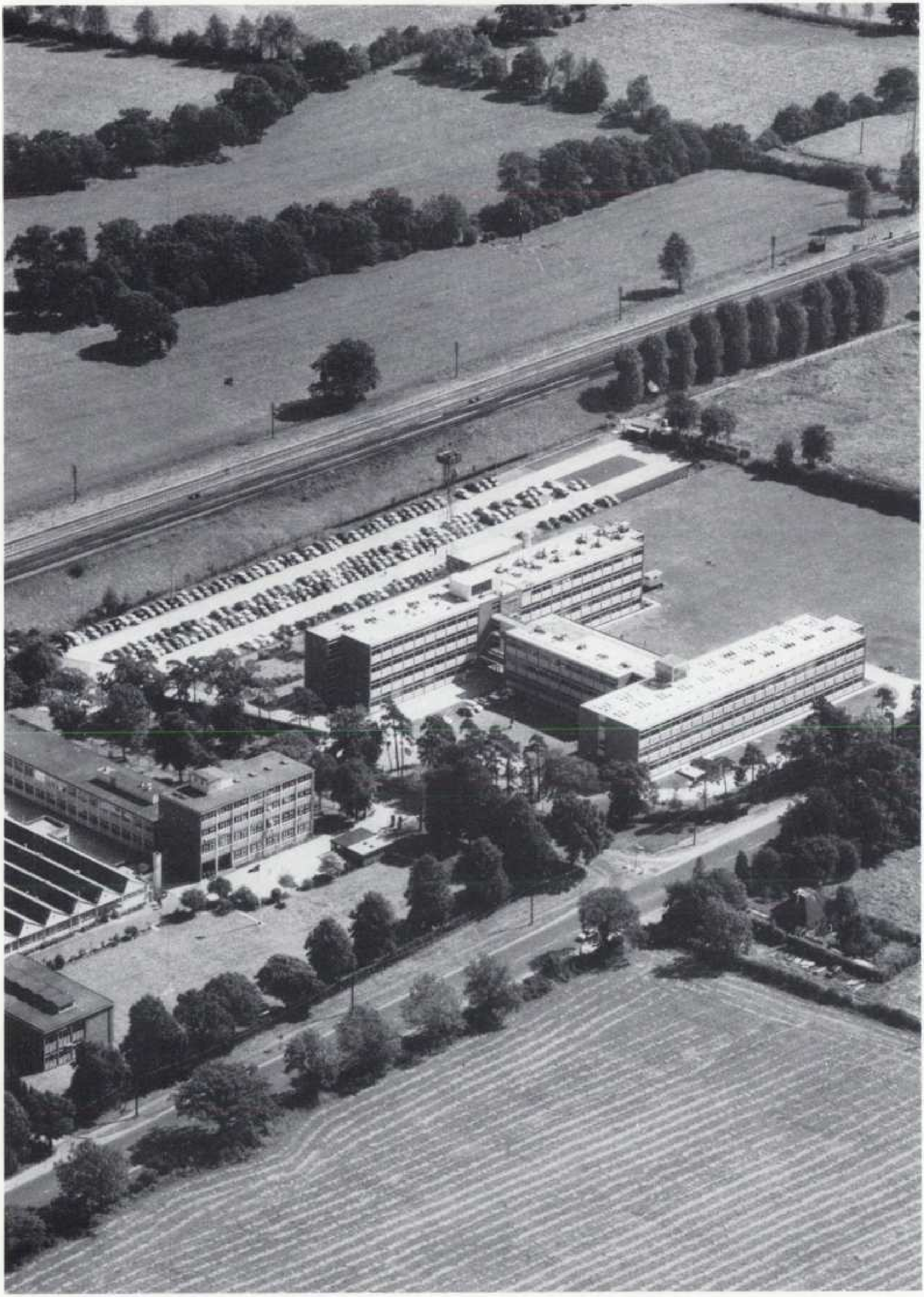
Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023


```
      CALL CLINEQ(MAT, XX, F, NEQ)

      DØ 8 N = 1,NEQ
    8 X(NC,N) = XX(N)

      DØ 9 N = 1,NEQ
      NØ = 2*N - 1
      EN = NØ*NØ - KK
      SUM = 0
      DØ 10 I = 1,NEQ
      L = 2*I - 1 + J
      EL = L*L
      SUM = SUM + L*X(NC,I)/(EL + EN)
   10 CØNTINUE
      Y(NC,N) = F1(N)*SUM
    9 CØNTINUE

    3 CØNTINUE

      DØ 11 N = 1,NCP
      REF(N) = Y(1,N) — Y(2,N)
      TRM(N) = Y(1,N) + Y(2,N)

      NØ = 2*N - 1
      NE = 2*N
      DEF(NØ) = X(1,N)
      DEF(NE) = - X(2,N)
   11 CØNTINUE
      TRM(1) = TRM(1) + ER(1)

      DØ 12 N = 1,NCP
      AREF(N) = CABS(REF(N))
      PREF(N) = CANG(REF(N))
      ATRM(N) = CABS(TRM(N))
      PTRM(N) = CANG(TRM(N))
   12 CØNTINUE
      DØ 13 N = 1,NEQ
      ADEF(N) = CABS(DEF(N))
      PDEF(N) = CANG(DEF(N))
   13 CØNTINUE

      SUM = 0
      DØ 14 N = 1,NCP
   14 SUM = SUM + AREF(N)*AREF(N)*RT(2*N - 1)
      EREF = SUM/RT(1)
      SUM = 0
      DØ 15 N = 1,NCP
   15 SUM = SUM + ATRM(N)*ATRM(N)*RT(2*N - 1)
      ETRM = SUM/RT(1)
      SUM = 0
      DØ 16 N = 1,NEQ
   16 SUM = SUM + ADEF(N)*ADEF(N)*RT(N)
      EDEF = SUM/RT(1)
      SUMS = EREF + ETRM + 2.*EDEF
      PRINT 100, K, EREF, ETRM, EDEF, SUMS
      PRINT 101
      ZERØ = 0.

      DØ 17 N = 1,NCP
      M = 2*N - 1
      PRINT 102, M, AREF(N), ATRM(N), ADEF(M),
           PREF(N), PTRM(N), PDEF(M)
      MM = M + 1
      PRINT 102, MM, ZERØ, ZERØ, ADEF(MM), ZERØ,
           ZERØ, PDEF(MM)
   17 CØNTINUE

    1 CØNTINUE
      STØP

  100 FØRMAT (1H1,//,10X,*K = *,F5.1,5X,*EREF = *,
                F9.6,5X,*ETRM = *,F9.6,5X,*EDEF = *,
                F9.6,5X,*SUMS = *,F9.6,//)
  101 FØRMAT (10X,*N*,10X,*AREF*,10X,*ATRM*,
                10X,*ADEF*,10X,*PREF*,10X,*PTRM
                *,10X,*PDEF*,/)
  102 FØRMAT (9X,I2,6(5X,F9.6))
      END


      SUBRØUTINE  CLINEQ(A,X,B,N)
      DIMENSIØN A(N,N),X(N),B(N)
      TYPE  CØMPLEX A,X,B,DIVR
      DØ 5 I = 1,N
      II = I + 1
      DØ 2 J = II,N
    2 A(I,J) = A(I,J)/A(I,I)
      B(I) = B(I)/A(I,I)
      A(I,I) = (1.,0.)
      DØ 4 K = II,N
      DIVR = A(K,I)
      DØ 3 L = I,N
    3 A(K,L) = A(K,L)/DIVR - A(I,L)
    4 B(K) = B(K)/DIVR - B(I)
    5 CØNTINUE
      MM = N + 1
    6 X(MM - 1) = B(MM - 1)
      DØ 7 M = MM,N
    7 X(MM - 1) = X(MM - 1) - X(M)*A(MM - 1,M)
      IF(MM.EQ.2) GØTØ 8
      MM = MM - 1
      GØTØ 6
    8 RETURN
      END
```

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023



Mullard Research Laboratories bij Redhill, Surrey, Engeland.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Filters, werkend met akoestische oppervlaktegolven

R. F. Mitchell

Inleiding

Kristaloscillators, filters en resonators zijn bekende voorbeelden van de sinds vele jaren gebruikelijke toepassing van hoogfrequente akoestische golven voor de bewerking van elektrische signalen. Over het geheel wordt hierbij gebruik gemaakt van scherpe resonanties, die optreden als gevolg van de herhaalde terugkaatsing van akoestische golven die zich voortplanten in het inwendige van regelmatige lichamen als platen, staven en schijven.

Sinds enige jaren is men tot het inzicht gekomen dat men voor vele vormen van signaalbewerking, waarbij scherpe resonanties niet noodzakelijk zijn, met voordeel gebruik kan maken van een ander soort akoestische evenwichtsverstoring: de akoestische oppervlaktegolf. De eigenschappen van deze golven zijn in vele centra onderzocht, waaronder Philips' Natuurkundig Laboratorium, Philips Forschungslaboratorium Hamburg en Mullard Research Laboratories (MRL). Akoestische oppervlaktegolven hebben met ruimtegolven vele eigenschappen gemeen — zo treedt er bij de voortplanting geen dispersie op —, maar verschillen daarvan doordat de evenwichtsverstoring beperkt is tot een klein gebied onder een oppervlak. De golven kunnen daardoor zonder bezwaar vanuit verscheidene punten gelijktijdig worden uitgezonden of daar worden ontvangen, en dit maakt een wijze van signaalverwerking mogelijk die sterk verschilt van de bij gangbare akoestische inrichtingen gebruikelijke.

Akoestische oppervlaktegolven hebben reeds op vele gebieden een nuttige toepassing gevonden, zoals in vertragingslijnen, impulscompressors, correlatie- en convolutieschakelingen voor signaalherkenning, maar de meestbelovende toepassing wordt gevonden in een bepaald soort eenvoudige analoge filters.

Wij zullen in dit artikel een kort overzicht geven van het theoretische en praktische werk dat gedurende de laatste twee, drie jaren in dit laboratorium aan het maken van zulke filters is besteed. De aandacht is hierbij in de eerste plaats gericht geweest op de mogelijkheid van het maken van een middenfrequent filter voor televisie dat geschikt is voor hybridische integratie; de verkregen resultaten worden verderop in dit artikel besproken. De beschreven technieken zijn echter bruikbaar voor een reeks bredebandfilters voor het frequentiegebied van 10 tot 1000 MHz, met bandbreedten van 100 kHz en meer. Wij zullen hier uitsluitend de algemene beginselen van filters met oppervlaktegolven beschrijven; verdere bijzonderheden, vooral de wiskundige methodes betreffende, zijn elders gepubliceerd [1]. Wij beginnen allereerst met een schets van de eigenschappen der akoestische oppervlaktegolven zelf.

Oppervlaktegolven in vaste lichamen

Golven aan vloeistofoppervlakken zijn voor ieder een vertrouwd natuurverschijnsel. Men is echter pas tot het inzicht gekomen dat dergelijke golven ook aan vaste lichamen kunnen optreden nadat Lord Rayleigh dit in 1885 wiskundig had voorspeld. De proefondervindelijke bevestiging dat zij als gevolg van aardbevingen optreden, kwam nog later. Deze golven, die naar hun ontdekker vaak Rayleigh-golven worden genoemd, planten zich veel sneller voort dan hun tegenhangers in vloeistoffen (gangbare snelheden zijn 1 tot 4 km/s) en gaan gepaard met veel kleinere oppervlaktebewegingen. De bewegingen die de deeltjes uitvoeren gelijken sterk op die welke in vloeistoffen optreden (*fig. 1*), met het kenmerkende verschil echter dat bij de Rayleigh-golf de elliptische beweging aan de oppervlakte tegengesteld gericht is aan de voortplantingsrichting, terwijl deze bij vloeistoffen daarmee gelijkgericht is. De verschillen tussen de twee soorten golven zijn een gevolg van het feit dat bij een vloeistofgolf de potentiële energie zijn oorsprong vindt in de zwaartekracht en bij een vast lichaam in de elasticiteit.

Bij een Rayleigh-golf neemt de evenwichtsverstoring exponentieel af wanneer men zich van de oppervlakte verwijdert, en wordt na twee à drie golflengten verwaarloosbaar klein. Dit betekent dat het voor praktische toepassingen voldoende is het substraat twee à drie golflengten dik te nemen om een voortplanting van Rayleigh-golven te verkrijgen die van voortplanting in een volledige halve ruimte niet te onderscheiden is.

Bij Rayleigh-golven in anisotrope media zijn de bewegingen der deeltjes ingewikkelder dan in fig. 1 is weergegeven, hoewel de hoofdkenmerken gewoonlijk zeer wel vergelijkbaar zijn. De normaal op het golffront en de energiestroom hebben echter in het algemeen niet precies dezelfde richting, zoals in het isotrope geval.

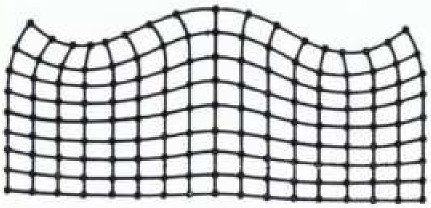
R. F. Mitchell, Ph.D., medewerker van Mullard Research Laboratories, Redhill, Surrey, Engeland.

[1] R. F. Mitchell, Generation and detection of sound by distributed piezoelectric sources, Philips Res. Repts. Suppl. 1972, no. 3.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Kenmerkende waarden voor de lengten van Rayleigh-golven zijn 1 mm bij 3 MHz, 10 μm bij 300 MHz en 1 μm bij 3 GHz, zodat te verwachten valt dat de afmetingen van onderdelen die met oppervlaktegolven werken van dezelfde orde van grootte zullen zijn. Zij zullen dus globaal 10^5 maal zo klein zijn als hun elektromagnetische tegenhangers.



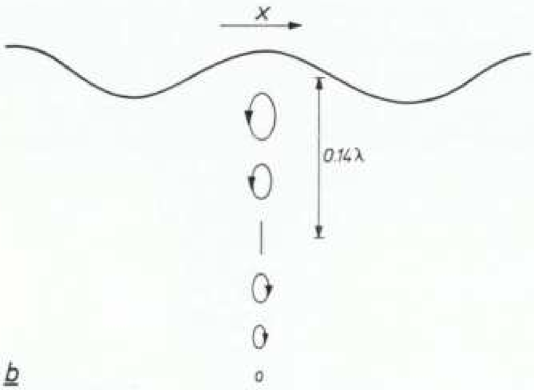


Fig. 1. *a*) Ter illustratie van de verplaatsingen die een akoestische oppervlaktegolf (een Rayleigh-golf) bij voortplanting in een isotroop medium teweegbrengt. Deze doorsnede geeft de verplaatsingen weer van deeltjes die zich oorspronkelijk op de snijpunten van een rechthoekig „rooster" bevonden. Bij een aanstootfrequentie van 30 MHz is de golflengte λ ongeveer 0,1 mm. *b*) Het elliptische karakter van de deeltjesbeweging in een isotroop medium. Aan het oppervlak doorloopt het deeltje zijn ellips in een richting, tegengesteld aan de richting van voortplanting x. De elliptische beweging keert van richting om op $0{,}14\lambda$ onder het oppervlak, en de verstoring wordt verwaarloosbaar klein binnen een afstand van ongeveer 3λ van het oppervlak.

Analoge filters met oppervlaktegolven

Algemene gezichtspunten

Akoestische oppervlaktegolven kunnen op eenvoudige wijze worden opgewekt en gedetecteerd met behulp van een interdigitale omzettersconfiguratie — dat is een configuratie, bestaande uit twee kammen waarvan de tanden in elkaar grijpen — op een sterk piëzo-elektrisch substraat (zie *fig. 2*). De metalen elektrodes worden gewoonlijk enige golflengten lang gemaakt, om er zeker van te zijn dat de uitgezonden golf „vlak" is (om precies te zijn: een recht golffront heeft), en liggen bij de grondfrequentie een halve golflengte uiteen. Dit betekent dat bij deze afmetingen de elektrodes voor frequenties van enige honderden MHz kunnen worden gemaakt met behulp van gangbare fotolithografische methodes, wanneer althans het oppervlak van het substraat een juiste voorbehandeling heeft ondergaan.

Fig. 3 is een schematische voorstelling van een eenvoudige vertragingslijn met een zendende en een ontvangende omzetter. Dit is de grondslag voor een filter werkend met oppervlaktegolven. De filterwerking berust op het feit dat het gedrag, zowel van de zendende als van de ontvangende omzetter, sterk van de frequentie afhangt. De frequentiekarakteristiek van het gehele filter is niets anders dan het produkt van de frequentiekarakteristieken van de twee omzetters.

Uit het schema van fig. 3 zijn veel van de aantrekkelijke eigenschappen van oppervlaktegolffilters af te leiden:

1. Hun geringe afmetingen: kenmerkende waarden zijn enkele tientallen akoestische golflengten.
2. Hun gemakkelijke vervaardiging: al wat men nodig heeft is een passend piëzo-elektrisch substraat waarvan één oppervlak tot op ongeveer een akoestische golflengte vlak is gepolijst. Zolang het substraat meer dan een paar golflengten dik is en groot genoeg om het filter erop te kunnen onderbrengen, zijn zijn overige afmetingen niet van groot belang. Verder kan het worden gemonteerd — bij voorbeeld door het in een houder te lijmen — zonder dat dit de werking beïnvloedt.
3. Hun eenvoud: het enige deel van het filter dat een hoge maatnauwkeurigheid eist is de elektrodenconfiguratie, en deze kan met bestaande middelen worden vervaardigd.

De gemakkelijkste manier om het gedrag van dit soort filters te verklaren is de uiteenzetting van de theorie in twee stappen te geven: de eerste is een eenvoudige analyse, waarbij een aantal benaderingen wordt gemaakt en die de algemene beginselen van het gedrag duidelijk doet uitkomen; bij de tweede gaat de behandeling meer in bijzonderheden en worden enkele der secundaire effecten besproken. Oppervlaktegolven zijn ingewikkelde verschijnselen en zelfs simpele problemen als de terugkaatsing van een oppervlaktegolf aan een hoek zijn slechts onder zeer eenvoudige voorwaarden voor wiskundige behandeling toegankelijk. Wanneer men elastische en piëzo-elektrische anisotropie invoert, tegelijk met periodieke belasting van het oppervlak door de elektrodenconfiguratie, worden de problemen formidabel van omvang. Het is dan ook nog niet mogelijk geweest om een volledige analyse van het gedrag van een eindige interdigitale omzetter te geven zonder benaderingen in te voeren. Wij zullen daarom gebruik maken van een aantal benaderingen en van analogieredeneringen om tot een bevredigend model van de situatie te komen.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

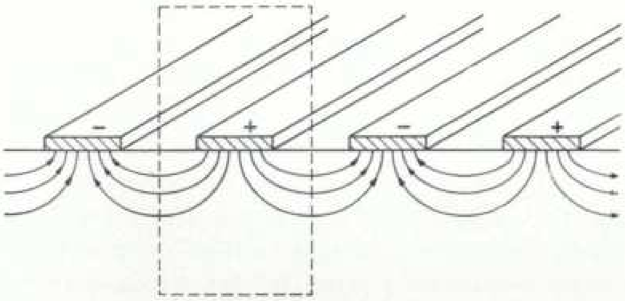


Fig. 2. Interdigitale omzetterconfiguratie voor het opwekken en detecteren van akoestische oppervlaktegolven. De metalen elektrodes, die op een sterk piëzo-elektrisch substraat zijn neergeslagen, zijn gewoonlijk verscheidene golflengten lang; bij de werkfrequentie is hun onderlinge afstand een halve golflengte. De elektrodes worden om en om in tegengestelde fase aangestoten. Deze configuratie kan het eenvoudigst worden gemaakt met behulp van fotolithografische methodes.

Binnen de gestippelde rechthoek bevindt zich de doorsnede van de elementaire eenheid, die later in bijzonderheden wordt geanalyseerd.

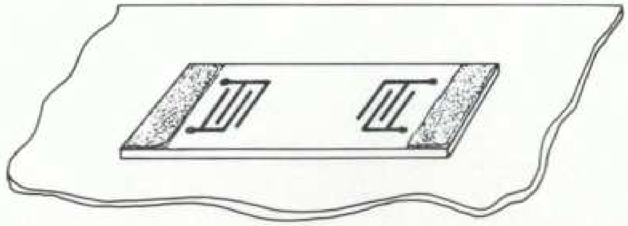


Fig. 3. Schematische tekening van een oppervlaktegolffilter met een zendende en een ontvangende omzetter. De filterwerking van de opstelling berust op de sterke frequentieafhankelijkheid van de twee omzetters. Zijn frequentiekarakteristiek is niets anders dan het produkt van de twee frequentiekarakteristieken. Het mechanische ontwerp is eenvoudig en het filter is zeer klein; het meet meestal niet meer dan enige tientallen akoestische golflengten. De geschaduwde delen aan de einden stellen absorberend materiaal voor dat ongewenste reflecties moet onderdrukken.

Vereenvoudigd model

De eenvoudigste en meest doeltreffende benadering is om elke elektrode van de omzetter als een enkelvoudige bron van oppervlaktegolven te beschouwen. Hierbij worden de volgende veronderstellingen gemaakt:

1. Alle enkelvoudige bronnen genereren golven die bij alle frequenties dezelfde amplitude hebben — d.w.z. deze bronnen hebben een vlakke frequentiekarakteristiek.
2. Er is geen akoestische demping.
3. De bronnen zijn onafhankelijk van elkaar.
4. Er zijn geen buigingsverschijnselen; er worden uitsluitend „vlakke" golven uitgezonden.

Onder deze omstandigheden is de omzetter niet anders dan een eenvoudige, aan één zijde aangesloten akoestische antenne, en de uitgezonden golf is de som van de bijdragen der individuele elektrodes. Voor sommige frequenties ondersteunen deze samenstellende golven elkaar en voor andere doven ze elkaar uit, zodat de opstelling een geprononceerde selectiviteit vertoont.

Een eenvoudige wiskundige analyse brengt de belangrijkste gedragskenmerken aan het licht.

We stellen de amplitude van de in het punt x_0 gegenereerde oppervlaktegolf voor door $B(x_0)$ en nemen aan dat de golf zich langs het oppervlak in de x-richting voortplant, loodrecht op de elektrodes. Aangezien de indringdiepte van de golf in het substraat constant is, mogen we dan voor de amplitude van de resulterende golf die een punt x buiten de omzetter bereikt, schrijven:

$$A(x) = c \int_a^b B(x_0) \exp\,[\mathrm{j}\{\omega t - k(x - x_0)\}]\,\mathrm{d}x_0,$$

waarbij de omzetter zich van $x_0 = a$ tot $x_0 = b$ uitstrekt; c is een constante. Aangezien $B(x_0)$ buiten ab nul is, mogen de grenzen tot $\pm\,\infty$ worden verlegd, hetgeen leidt tot:

$$A(x) = c\,.\,\mathcal{F}(B) \exp \mathrm{j}(\omega t - kx),$$

waarin

$$\mathcal{F}(B) = \int_{-\infty}^{\infty} B(x_0) \exp\,(\mathrm{j}kx_0)\,\mathrm{d}x_0$$

de Fourier-getransformeerde van B is. Wij zien dus dat er een lopende oppervlaktegolf wordt uitgezonden waarvan de amplitude de Fourier-getransformeerde $\mathcal{F}(B)$ is van de bronsterkteverdeling, $B(x_0)$. De functie $\mathcal{F}(B)$ is in het algemeen complex en geeft zowel de amplitude als de fase van de golf. Uit deze eenvoudige betrekking kan gemakkelijk worden afgeleid hoe de doorlaatband van een filter samenhangt met de elektrodenconfiguratie. Wij beschouwen allereerst het geval van een eenvoudige omzetter met uniforme configuratie die, volgens dit eenvoudige model, kan worden voorgesteld als een reeks enkelvoudige bronnen, één voor elke elektrode (*fig. 4*). Het teken bij elke bron wisselt met de aan de elektrodes aangelegde span-

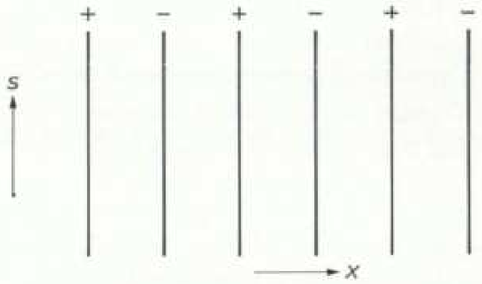


Fig. 4. Een eenvoudige uniforme omzetter kan worden voorgesteld door een reeks enkelvoudige bronnen, één bron voor elke elektrode. De tekens der bronnen wisselen met de spanning die aan de elektrodes wordt aangelegd. De sterkte s van elke bron komt in de lengte van de lijn tot uitdrukking. In dit geval hebben de bronnen alle dezelfde sterkte en de omhullende kromme is een rechthoekige functie van de lineaire coördinaat x.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

ning. De Fourier-getransformeerde van de resulterende bronverdeling kan op eenvoudige wijze worden gevonden door op te merken dat de verdeling gelijkwaardig is met een rechthoekige omhullende functie die op regelmatige afstanden wordt afgetast. Het is bekend dat de Fourier-getransformeerde van een dergelijke functie de getransformeerde is van de omhullende curve, herhaald op harmonische intervallen waarvan de grootte wordt bepaald door de onderlinge afstand der aftastpunten. Aangezien de Fourier-getransformeerde van een rechthoekige functie van de vorm $(\sin x)/x$ is, ziet de getransformeerde van de functie van fig. 4 er uit als is weergegeven in *fig. 5*. Dit is dus de verwachte frequentiekarakteristiek van een uniforme omzetter. Deze verwachting wordt in hoofdlijnen door het experiment bevestigd.

De eenvoudige betrekking tussen de omhullende van de bronsterkteverdeling en de frequentiekarakteristiek van de omzetter maakt het gemakkelijk, te besluiten hoe de bronsterkteverdeling gewijzigd moet worden om een bepaalde frequentiekarakteristiek te verkrijgen. Wanneer de bandbreedte van de omzetter maar niet te groot is, interfereren de karakteristieken voor de harmonischen niet met elkaar en kan de karakteristiek voor de grondgolf op zichzelf worden beschouwd. *Fig. 6* geeft een aantal bronsterkteverdelingen en de daarmee overeenkomende karakteristieken voor het frequentiegebied rondom de grondfrequentie, berekend uit de getransformeerde van de omhullende.

Mogelijke doorlaatkrommen

Uit het door deze eenvoudige analyse verkregen resultaat in de vorm van Fourier-getransformeerden kunnen de volgende algemene beginselen worden afgeleid, zolang althans onze veronderstelling geldt dat de onderlinge afstand der elektrodes gelijk is.

1. De werkfrequentie hangt af van de afstand der elektrodes; hoe kleiner deze is, hoe hoger de frequentie. Wij hebben er reeds op gewezen dat de frequentie van de eerste harmonische die is waarvoor de elektrodenafstand gelijk is aan de halve golflengte.

2. De relatieve bandbreedte voor de grondgolf is omgekeerd evenredig met het aantal bronnen; hoe meer bronnen er zijn, hoe kleiner de bandbreedte. Aangezien de centrale frequentie afhangt van de afstand der elektrodes en van de voortplantingssnelheid der oppervlaktegolven, volgt hieruit dat de absolute bandbreedte van het filter uitsluitend afhangt van zijn grootte (in de richting van voortplanting) en van de voortplantingssnelheid. Als de snelheid b.v. 3000 m/s is, zal de bandbreedte van een opstelling van een centimeter lang ongeveer 300 kHz zijn, van een opstelling van twee cm lang 150 kHz, enz. Deze betrekking tussen bandbreedte en afmeting vormt een van de belangrijkste nadelen van oppervlaktegolffilters, daar filters met smalle doorlaatband zeer grote substraten vereisen. Dit probleem zou zijn opgelost indien wij beschikten over een gemakkelijke en efficiënte methode om oppervlaktegolven te laten reflecteren, wat nog niet het geval is.

3. De Fourier-getransformeerde van een symmetrische functie is reëel; die van een asymmetrische functie is zuiver imaginair. Elke symmetrische of antisymmetrische bronsterkteverdeling zal dus een signaal genereren waarvoor de relatie tussen fase en frequentie lineair is. Een groot aantal amplitudekarakteristieken is mogelijk bij gelijkblijvende fase omdat oppervlaktegolffilters geen minimumfase-netwerken zijn; amplitudevariaties hebben niet noodzakelijkerwijs fasevariaties tot gevolg. Om een niet-lineair faseverloop te verkrijgen, moeten onsymmetrische elektrodeconfiguraties

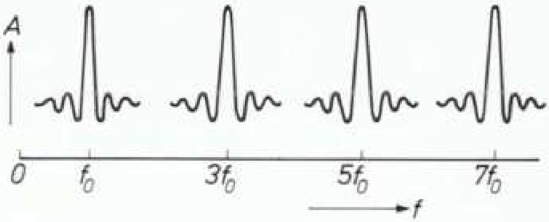


Fig. 5. Voor het eenvoudige omzettermodel is de voorspelde frequentiekarakteristiek evenredig met de modulus van de Fourier-getransformeerde van de omhullende kromme en herhaalt zich op een harmonisch interval waarvan de grootte wordt bepaald door de onderlinge afstand der elementen. Aangezien de Fourier-getransformeerde van een rechthoekige functie de vorm $(\sin x)/x$ heeft, ziet de getransformeerde van de functie van fig. 4, en dus de totale frequentiekarakteristiek, eruit zoals hier is weergegeven. (Merk op dat dit schema het reële deel van de Fourier-getransformeerde weergeeft.) A amplitude, f frequentie.

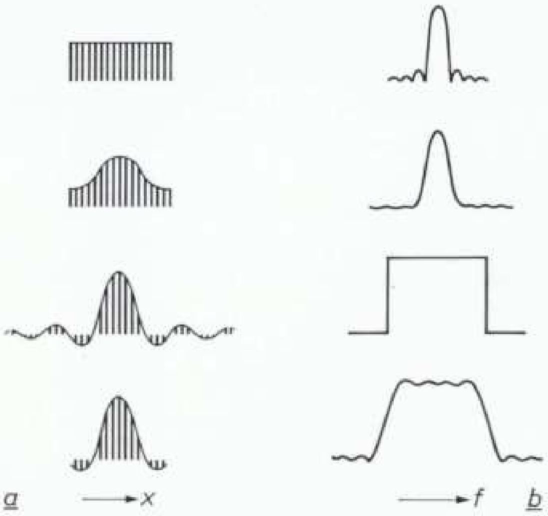


Fig. 6. Enkele bronsterkteverdelingen (*a*) met Fourier-getransformeerden (hier als de modulus weergegeven) van hun omhullenden (*b*). De Fourier-getransformeerden geven de karakteristiek voor frequenties rondom de grondfrequentie. Omtrent de bronverdelingen is weer ondersteld dat de elementen om en om verschillend teken hebben, en ook van teken wisselen als de omhullenden overgaan van positief in negatief.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

of veranderlijke elektrodenafstanden worden toegepast.
4. De Fourier-getransformeerde van een reële functie is hermitisch [*]. Dit betekent dat, binnen de beperkingen van het model, de frequentiekarakteristieken uitsluitend symmetrisch kunnen zijn. Afwijkingen hiervan maken het eenvoudige model ongeldig.

Synthese van een frequentiedoorlaatkromme

De benadering door middel van de Fourier-transformatie geeft een nuttige aanwijzing omtrent de frequentiedoorlaatkrommen die fysisch mogelijk zijn. Theoretisch zou dezelfde methode kunnen worden gebruikt om uit te maken welke omhullende van de bronsterkteverdeling vereist zou zijn om een bepaalde doorlaatkromme te verkrijgen, door daarvan de omgekeerde transformatie te nemen. De moeilijkheid is dat de Fourier-getransformeerde van de meeste functies zich tot plus en min oneindig uitstrekt; dit betekent dat de bronsterkteverdeling, vereist om precies een bepaalde verlangde doorlaatkromme te verkrijgen, in het algemeen een groot en wellicht oneindig aantal bronnen zal omvatten — een wat lastig te fabriceren configuratie. Voor de praktische uitvoering moet de bronsterkteverdeling na een klein aantal termen worden afgebroken, maar zo dat de afwijking van de verlangde kromme zo gering mogelijk is.

Dit is een moeilijk probleem, dat een studie op zichzelf vereist. Het blijkt het beste te zijn de frequentiekarakteristiek van een bronnenverdeling numeriek te berekenen zonder expliciet transformatieanalyse toe te passen en dan het verkregen resultaat te optimaliseren. Door list en zorgvuldigheid kan men een bronsterkteverdeling vinden die de vereiste kromme oplevert, ook wanneer deze asymmetrisch moet zijn en de relatie tussen fase en frequentie niet-lineair. In ons laboratorium wordt de optimalisatie verkregen door een „wisselwerking tussen mens en machine", hoewel er thans een rekenprogramma geschreven is dat, voor het geval van een lineaire faserelatie, automatisch voor een verlangde doorlaatkromme de vereiste bronsterkten berekent.

De verdeling der bronsterkten

Bij de bespreking van het ontwerp van oppervlaktegolffilters is tot nu toe uitsluitend de bronsterkteverdeling ter sprake gekomen en zijn de methodes die praktisch bruikbaar zijn ter verkrijging van de vereiste bronsterkten, buiten beschouwing gebleven. Ter beinvloeding van de sterkteverdeling van de bronnen kan men twee hoofdwegen bewandelen: het variëren van de breedte der elektrodes (*fig. 7*) en het variëren van hun lengte (*fig. 8*). De methode der breedtevariatie heeft het grote voordeel dat buigingseffecten daardoor tot een minimum worden teruggebracht. Het nadeel is echter dat voor bronnen van geringe sterkte zeer smalle elektrodes nodig zijn, waarvoor de fotolithografische techniek dan weer overeenkomstig nauwkeurig moet zijn. Dit betekent dat variatie van de breedte slechts bij lage frequenties kan worden gebruikt; bij 30 MHz kan men b.v. met behulp van de gangbare lithografische technieken geen grotere variatie in bronsterkte verkrijgen dan in een verhouding van 2,5 op 1. De sterktebeïnvloeding door lengtevariatie stelt minder eisen aan de nauwkeurigheid van fabricage, maar doet de buigingseffecten toenemen. Deze effecten doen zich natuurlijk het sterkst gelden in filters waarvan de „opening" — d.i. de lengte van de langste „tand" van de kamvormige elektrode, uitgedrukt in golflengten — klein is; hier draagt de diffractie een hele reeks nieuwe ont-

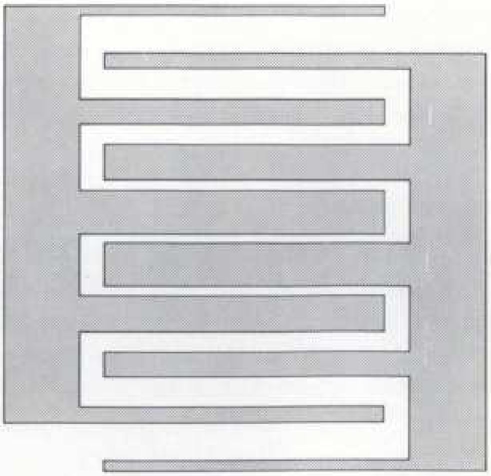


Fig. 7. Een manier om tot een passende bronsterkteverdeling te komen is variatie van de breedte der „tanden" die de elementen vormen. Deze methode is slechts bruikbaar bij filters voor lage frequenties. Hij heeft het voordeel dat er slechts weinig buiging bij optreedt.

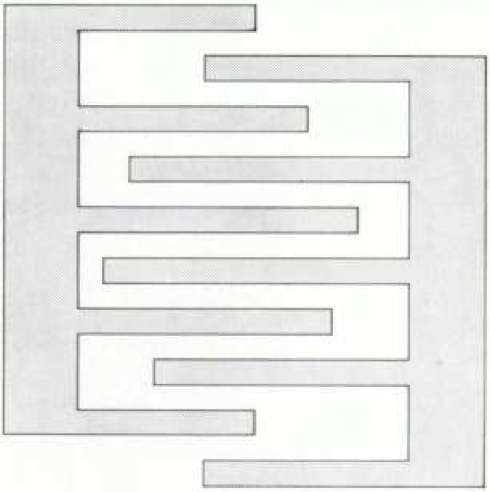


Fig. 8. De bronsterkteverdeling kan ook worden beïnvloed door lengtevariatie van de tanden.

[*] Van een hermitische functie is het reële deel symmetrisch om de oorsprong, terwijl het imaginaire deel antisymmetrisch is. De modulus van een dergelijke functie is symmetrisch.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

werpproblemen aan. Dit is de reden dat wij bij ons werk aan middenfrequent-televisiefilters de meeste aandacht aan sterktebeïnvloeding door breedtevariatie hebben besteed; in de naaste toekomst zijn wij echter van plan om op lengtevariatie over te gaan.

Het volledige model

Bij het ontwerpen van een filter kan men het best van het eenvoudige model uitgaan, maar de eigenschappen van het filter dat men verkrijgt zullen wat anders zijn dan het model aangeeft, omdat dan de secundaire effecten om de hoek komen kijken die tot nu toe buiten beschouwing zijn gebleven. *Fig. 9* geeft een vergelijking tussen de werkelijke eigenschappen van een filter met door breedtevariatie verkregen bronsterkteverdeling, en de op grond van het eenvoudige model voorspelde. De overeenstemming is heel behoorlijk, maar er zijn drie belangrijke discrepanties: de diepten der minima zijn geheel verkeerd voorspeld, de kromme vertoont in het middendeel een scheefheid waarvan de theorie geen rekenschap geeft, en in het gebied van de zijlobben wordt de voorspelling in toenemende mate onbetrouwbaar. Het is van belang het eenvoudige model te verfijnen, zodat met deze afwijkingen rekening kan worden gehouden en ze bij het ontwerp kunnen worden weggewerkt voordat de filters gemaakt worden. Er is dus een ingewikkelder model nodig, hetgeen vereist dat sommige der voor het eenvoudige model gemaakte veronderstellingen verlaten worden. Dit volledige model moet rekening houden met de effecten van buiging, akoestische demping, wisselwerkingen tussen de individuele bronnen en variatie van het per bron afgegeven vermogen met de frequentie. Wij zullen thans in het kort bespreken hoe wij het probleem van het voorspellen van deze effecten hebben aangepakt.

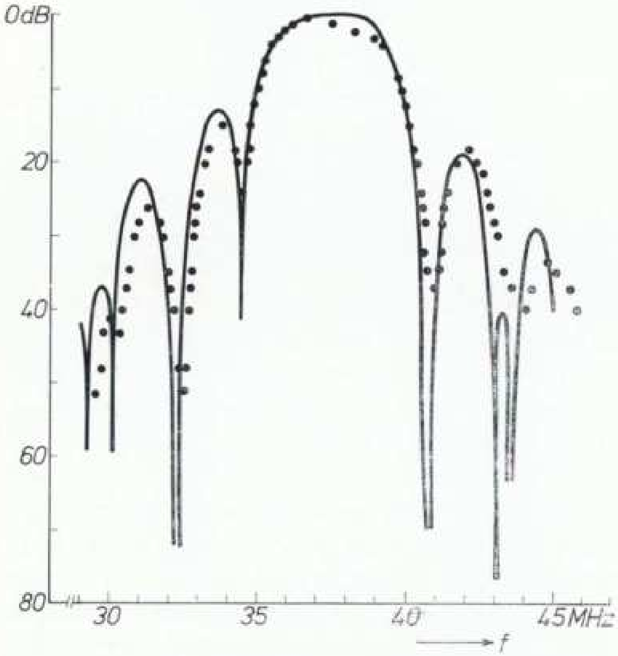


Fig. 9. Experimentele resultaten (stippen) voor de doorlaatkromme van een filter waarbij de bronsterkteverdeling door breedtevariatie is verkregen, vergeleken met de berekening volgens de eenvoudige theorie (getrokken kromme). De overeenstemming is zeer behoorlijk, doch er zijn drie belangrijke discrepanties: de minima zijn te ondiep, de doorlaatkromme heeft een onverwachte scheefheid in het middendeel, en de voorspelling wordt steeds onbetrouwbaarder naarmate men zich verder in het gebied der zijlobben begeeft.

De frequentiekarakteristiek van elke bron afzonderlijk

Om voor elke elektrode afzonderlijk de frequentiekarakteristiek te bepalen maken wij gebruik van een modificatie van een berekening van Engan [2]. Hierbij wordt het probleem benaderd door het elektrische veld dat door een interdigitaal elektrodensysteem wordt opgewekt, te ontwikkelen in een Fourier-reeks waarvan de termen, evenals bij een Rayleigh-golf, langs het oppervlak sinusvormig variëren, maar daaronder exponentieel afnemen. De opwekking van oppervlaktegolven bij een bepaalde frequentie is verbonden met de component van het elektrische veld die de overeenkomstige ruimtelijke periodiciteit heeft. Zo is een oppervlaktegolf met golflengte λ verbonden met de component van het elektrische veld waarvan de herhalingsafstand λ is. Er wordt aangenomen die de amplitude van de opgewekte akoestische golf evenredig is met de amplitude van de overeenkomstige component van het elektrische veld. De door Engan oorspronkelijk gemaakte analyse gold voor een oneindig systeem van interdigitale elektrodes. Wij hebben deze analyse aangepast aan het geval van een enkele elektrode door een gelijksoortige reeksontwikkeling op te schrijven voor de zich herhalende eenheid die in fig. 2 door een rechthoek is omgeven — een enkele elektrode die deel uitmaakt van een oneindig systeem. Een omzetter die elektrodes van veranderlijke breedte bevat, kan bij benadering worden voorgesteld door herhalingseenheden van dit type aaneen te voegen.

Fig. 10 geeft de op deze wijze bepaalde frequentiekarakteristieken voor twee op zichzelf beschouwde eenheden uit elektrodensystemen met verschillende verhoudingen tussen breedte en tussenruimte der elektrodes. De amplitude wijzigt zich niet snel met de frequentie, en als men dus de afmetingen der elektrodes zo kiest dat men de verlangde bronsterkten bij de grondfrequentie verkrijgt, blijven de gekozen waarden ongeveer juist in een band van aanzienlijke breedte. In dit gebied is echter de variatie voor brede en smalle elektrodes tegengesteld, waardoor de uitwerking op de frequentiekarakteristiek van een filter waarbij de sterkteverdeling door breedtevariatie is verkregen, en dat dus beide soorten elektrodes bevat, aanzienlijk kan zijn. Komt men

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

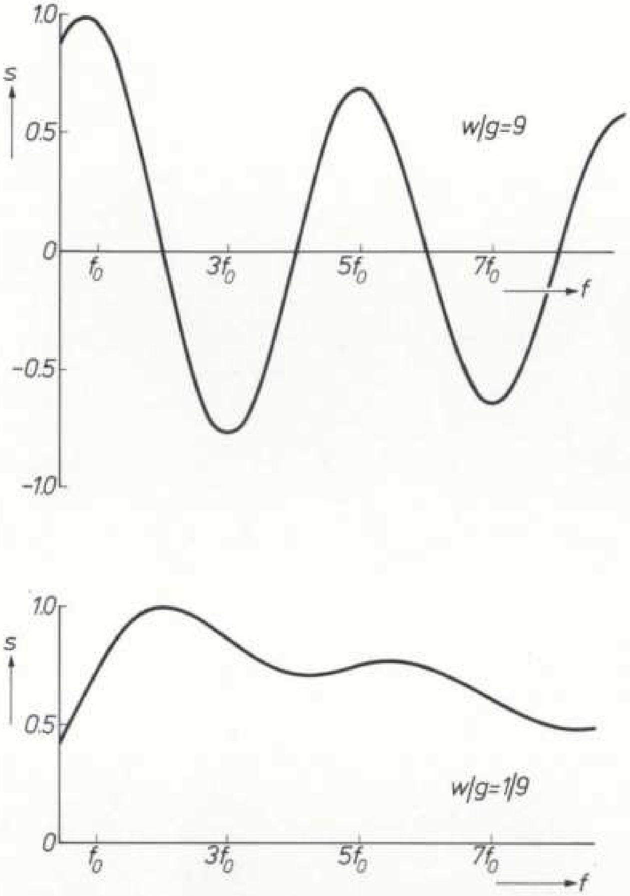


Fig. 10. Frequentiekarakteristieken voor twee op zichzelf beschouwde eenheden, bestaande uit kammen met verschillende verhouding w/g tussen breedte en tussenruimte der elektrodes. s amplitude van de oppervlaktegolf. De aanslag wijzigt zich niet snel met de frequenties, zodat, wanneer men de afmetingen der elektrodes zo kiest dat zij de verlangde bronsterkten leveren bij de grondfrequentie f_0, deze keuze over een band van aanzienlijke breedte juist zal blijven. Aangezien echter de aanslagvariatie in het gebied rondom f_0 voor brede en smalle elektrodes tegengesteld is, is de uitwerking op de karakteristiek van een omzetter waarbij de bronsterkteverdeling door breedtevariatie is verkregen, en die beide typen elektrodes bevat, aanzienlijk. Voor filters met lengtevariaties der elektrodes behoeft de uit het eenvoudige model voorspelde aanslag slechts te worden vermenigvuldigd met de aanslag van een enkele elektrode. Voor $w/g = 1$ is de aanslag op de derde harmonische nul.

ver boven de grondfrequentie, dan wordt door het verschil in de karakteristieken der elektrodes de verhouding der bronsterkten volledig anders: sommige elektrodes wekken in het geheel geen golven meer op en sommige doen het zelfs in tegenfase. Voor een omzetter waarbij de bronsterkteverdeling door breedtevariatie is verkregen, is dan ook de karakteristiek voor de derde harmonische een volledig andere dan voor de grondfrequentie. Bij omzetters waarbij van lengtevariatie gebruik is gemaakt, is de toestand eenvoudiger doordat de frequentiekarakteristiek voor alle elektrodes dezelfde is; men behoeft de uit het eenvoudige model berekende karakteristiek slechts te vermenigvuldigen met die van een enkele elektrode. Dit beïnvloedt de hoofddoorlaatband slechts weinig, maar heeft wel invloed op de onderlinge sterkte der harmonischen. Wanneer de elektrodes even breed zijn als de ruimte ertussen, wordt de omzetter door de derde harmonische in het geheel niet aangesproken.

Wisselwerking tussen bronnen

Er is tweeërlei wisselwerking tussen de bronnen mogelijk. Allereerst is er een zuiver akoestisch effect, waarbij een oppervlaktegolf van een elektrode gedeeltelijk door een daarnaast liggende andere wordt gereflecteerd als gevolg van de verandering in elektrische en mechanische omstandigheden die hij ontmoet wanneer hij zich van het vrije oppervlak naar het gemetalliseerde gebied beweegt. Ten tweede is er een specifiek piëzo-elektrisch effect, waarbij een golf die door een elektrode is opgewekt gedeeltelijk door een volgende in een elektrisch signaal wordt omgezet en dan opnieuw uitgestraald. Dit is ook weer een soort reflectie, die echter enigszins wordt beïnvloed door de uitwendige elektrische schakeling. Ten einde de wisselwerkingseffecten mede in rekening te brengen, hebben wij een idee dat door W. R. Smith c.s. [3] naar voren is gebracht, verder ontwikkeld. Volgens dit idee kan men, uitgaande van het feit dat de voortplanting van oppervlaktegolven in wezen een eendimensioneel gebeuren is (afgezien van de buiging breidt de golf zich uitsluitend in één richting van voortplanting uit), tot een model van een omzetter met een periodieke ruimtelijke structuur komen door analyse van het gedrag van een systeem van omzetters van zich ruimtelijk uitbreidende akoestische golven, welke omzetters akoestisch in serie, maar elektrisch parallel zijn geschakeld (*fig. 11*). Voor elke elektrode van het met oppervlaktegolven werkende apparaat is er in het model een omzetter, terwijl voor elk daarvan de piëzo-elektrische activiteit op een voor de opwekking van oppervlaktegolven passende waarde wordt ingesteld (deze is veel lager dan voor ruimtegolven).

Ten einde zeker te zijn dat het model wisselwerkingseffecten correct weergeeft, moet men het juiste type omzetter voor ruimtegolven kiezen. Op grond van proeven met eenvoudige modellen [4] hebben wij een plaatvormige piëzo-elektrische omzetter als basiseenheid gekozen en zijn daarmee tot het model van fig. 11 gekomen. Het is geldig zowel wanneer de bronsterkteverdeling van de omzetter door breedte- als wanneer deze door lengtevariatie is verkregen. De maat x_1 is

[2] H. Engan, Excitation of elastic surface waves by spatial harmonics of interdigital transducers, IEEE Trans. **ED-16**, 1014-1017, 1969.
[3] W. R. Smith, H. M. Gerard, J. H. Collins, T. M. Reeder en H. J. Shaw, Analysis of interdigital surface wave transducers by use of an equivalent circuit model, IEEE Trans. **MTT-17**, 856-864, 1969.
[4] R. F. Mitchell, W. Willis en M. Redwood, Electrode interactions in acoustic surface-wave transducers, Electronics Letters **5**, 456-457, 1969.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

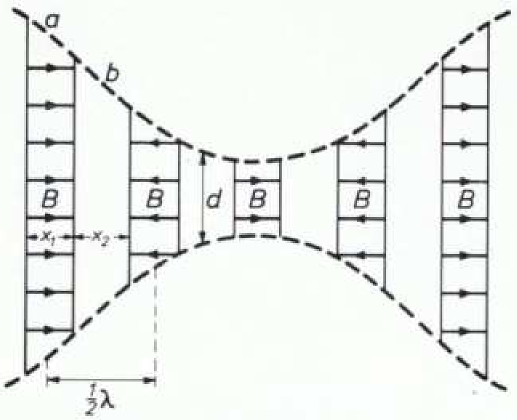


Fig. 11. Schema van de doorsnede van een model met akoestische massagolven (golflengte λ) als equivalent van een oppervlaktegolfomzetter met door lengte- of breedtevariatie verkregen bronsterkteverdeling. Dit model dient ter bestudering van wisselwerkingen tussen de bronnen. De streeplijnen stellen de boven- en ondervlakken voor van een akoestische golfgeleider. De ruimtegolfomzetters *B* (één voor elk element) staan akoestisch in serie en elektrisch parallel. De pijlen stellen het elektrische veld voor. De gebieden *a* en *b* zijn akoestisch aangepast en de verhouding x_1/x_2 is van geen belang. De afmeting *d* is zo gekozen dat de statische lading op elke omzetter dezelfde is als de lading op het overeenkomstige element van het oppervlaktegolffilter; de aan iedere elektrode aangelegde spanning is de helft van die van de overeenkomstige elektrode.

voor alle omzetters dezelfde en de afmeting loodrecht op het vlak van tekening is evenredig met de lengte van de overeenkomstige elektrode. Afmeting *d* is zo gekozen dat de statische lading die op elke omzetter van het model ontstaat, dezelfde is als die welke op de overeenkomstige elektrode van de oppervlaktegolfomzetter ontstaat. De omzetters zijn onderling verbonden door een akoestische golfgeleider die dezelfde werking heeft als het oppervlak, doordat hij ervoor zorgt dat alle akoestische energie die wordt opgewekt door een omzetter met grote *d*, ook door elke andere omzetter gaat, zelfs die met de kleinste waarde van *d*. Ten slotte wordt er met de frequentiekarakteristiek der individuele elektrodes rekening gehouden door de koppelingscoëfficiënt van elke omzetter zodanig met de frequentie te doen variëren dat het afgegeven akoestische vermogen op de juiste wijze van de frequentie afhangt (als in fig. 10).

Dit model wordt geanalyseerd door het oplossen van een reeks simultane vergelijkingen ter bepalen van de elektrische doorstroming door elke elektrode. De akoestische demping kan op eenvoudige wijze in rekening worden gebracht, evenals de buiging *binnen in* de omzetter, al lijkt dit op het ogenblik niet noodzakelijk te zijn.

Uit esthetisch oogpunt is dit model niet bijzonder fraai, maar men kan er op uitstekende wijze het gedrag van omzetters mee voorspellen, zoals te zien is uit *fig. 12*, die een vergelijking geeft tussen de theorie en de praktijk van een filter met door breedtevariatie ge-

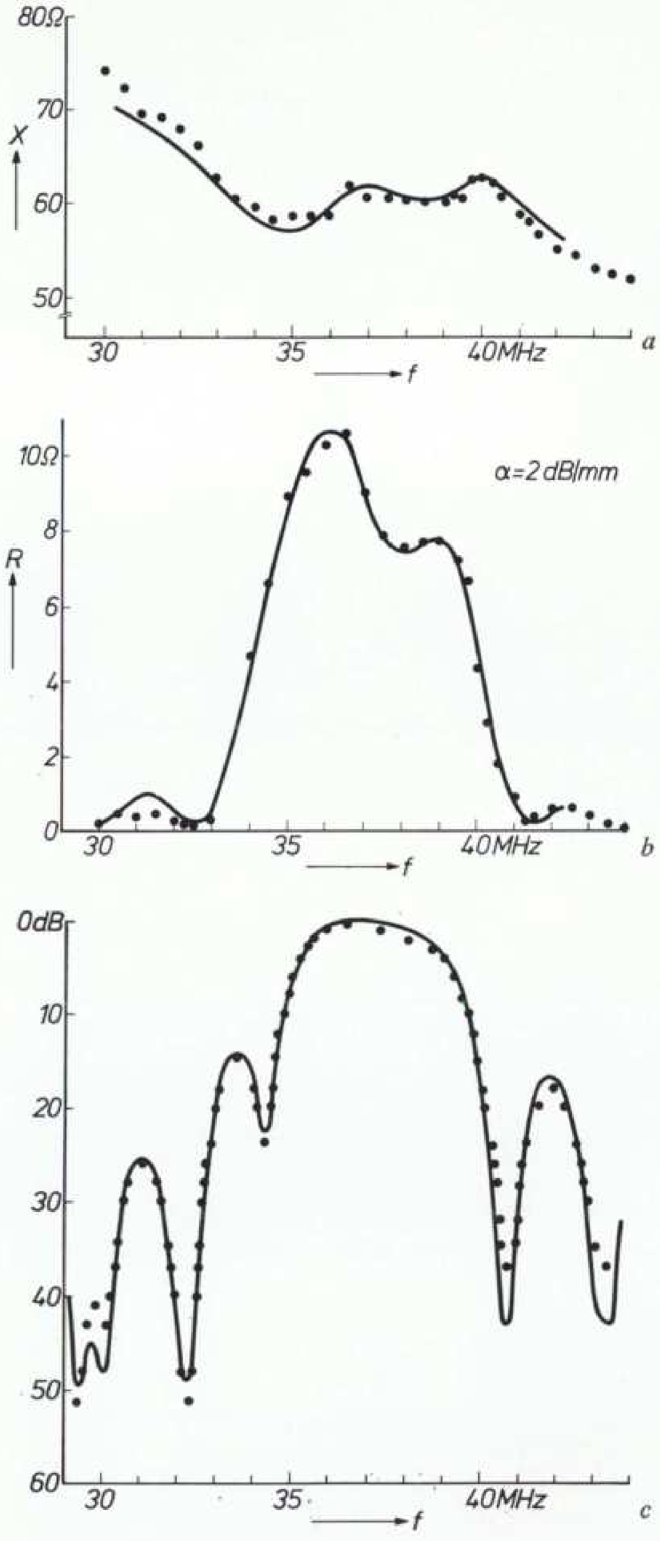


Fig. 12. Vergelijking van de volledige theorie (gebaseerd op het equivalente model met ruimtegolven; getrokken krommen) met experimentele resultaten (stippen) voor een filter met akoestische oppervlaktegolven waarbij de bronsterkteverdeling is verkregen door breedtevariatie. *a*) De reactantie *X* als functie van de frequentie. *b*) De weerstand *R* als functie van de frequentie; α is de demping. *c*) Doorlaatkromme van het filter. De overeenstemming is beter dan bij het eenvoudige model van de figuren 4 en 5. De scheefheid in het midden blijkt het gevolg te zijn van wisselwerkingseffecten, de diepten der minima worden door de demping bepaald en de amplituden der zijlobben zijn gewijzigd door de frequentiekarakteristiek der individuele elektrodes.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

regelde bronsterkteverdeling. De tekortkomingen van het eenvoudige model zijn nu opgeheven. De scheefheid in de doorlaatband blijkt het gevolg te zijn van wisselwerkingseffecten, de diepte der doorlaatminima wordt door de demping bepaald en de hoogte der zijlobben is gewijzigd via de frequentiekarakteristiek der individuele elektrodes.

Deze overeenstemming tussen theorie en praktijk is ruimschoots voldoende voor het maken van een ontwerp van filters met door breedtevariatie verkregen bronsterkteverdeling, maar voordat wij hiervan enkele toepassingen beschrijven moeten wij een andere groep secundaire effecten noemen die onder sommige omstandigheden belangrijk kunnen zijn.

Buigingseffecten

Voor filters met een kleine opening, in het bijzonder wanneer hun bronsterkteverdeling door lengtevariatie is verkregen, moeten de buigingseffecten in aanmerking worden genomen en zo nodig gecompenseerd. Het werk aan dit probleem is nog niet voltooid: wij kunnen theoretische voorspellingen maken maar hebben nog geen proefondervindelijke resultaten. Van het soort effecten dat een rol speelt krijgt men een indruk uit *fig. 13*, die het berekende buigingspatroon weergeeft van de oppervlaktegolven, opgewekt door een omzetter met 20 elektrodes en een opening van 100 golflengten. Dit is het „nabije" gebied van buiging, het gebied waarin de ontvangende omzetter van het filter gewoonlijk wordt ge-

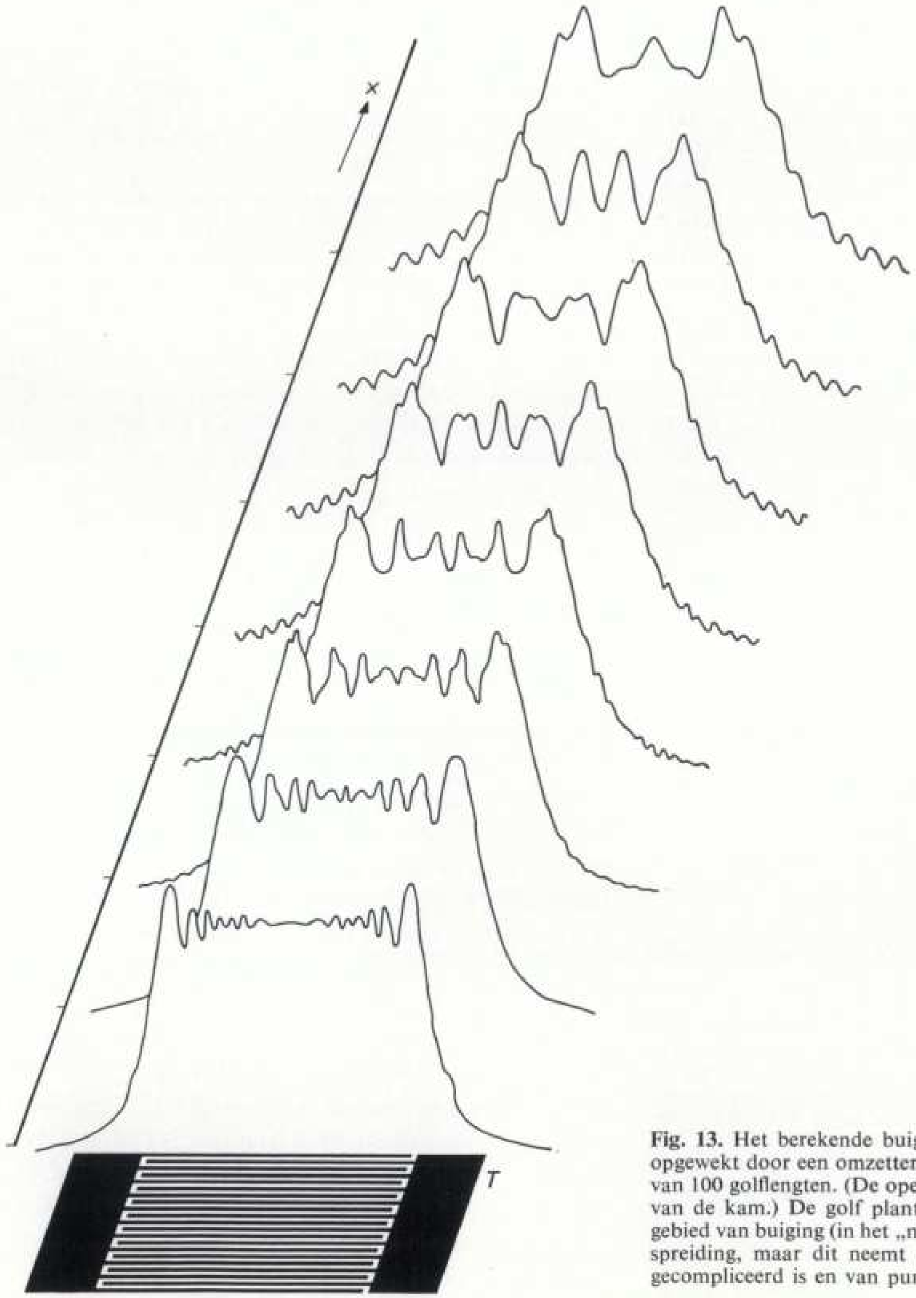


Fig. 13. Het berekende buigingspatroon der oppervlaktegolven opgewekt door een omzetter T met 20 elektrodes en een opening van 100 golflengten. (De opening is de lengte van de langste tand van de kam.) De golf plant zich voort in de x-richting. In dit gebied van buiging (in het „nabije veld") is er nog geen zijdelingse spreiding, maar dit neemt niet weg dat het buigingsveld zeer gecompliceerd is en van punt tot punt sterk verandert.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

plaatst. De golf is nog niet begonnen met zich zijdelings uit te breiden, maar het buigingsveld is desalniettemin zeer ingewikkeld en varieert ruimtelijk zeer sterk. Het is duidelijk dat de amplitude van de oppervlaktegolf die door een ontvangende elektrode in dit „nabije" veld wordt gedetecteerd, op zeer complexe wijze afhangt van de lengte van die elektrode, evenals van de plaats waar hij zich precies bevindt, en zeer sterk met de frequentie kan variëren.

In *fig. 14* zijn krommen uitgezet van de amplitudefouten door buiging die men aantreft in het signaal dat ontvangen wordt door een enkel ontvangend element, op verschillende afstanden van een zendende omzetter met 14 elektrodes van 10 golflengten lang. Het is belangwekkend dat voor een reeks afstanden de buiging het signaal groter maakt wanneer de ontvangende elektrode klein is. Heeft de elektrode dezelfde lengte als de tanden, dan zijn de fouten betrekkelijk klein; dit is de reden waarom de benadering via breedtevariatie voordelig is. Diffractiefouten variëren uit de aard der zaak met de frequentie en met de afstand, en zijn van bijzonder belang bij minima, waar geëist wordt dat alle individuele bijdragen elkaar nauwkeurig zullen uitdoven. Bij MRL beschikken wij over een rekenprogramma dat bij de centrale frequentie voor elke elektrode de door buiging veroorzaakte amplitude- en fasefouten bepaalt en daarna de lengte en de plaats der elektrodes wijzigt om de fouten te compenseren. De berekening toont aan dat deze, voor één frequentie uitgevoerde, compensatie over een band van aanzienlijke breedte redelijk van kracht blijft.

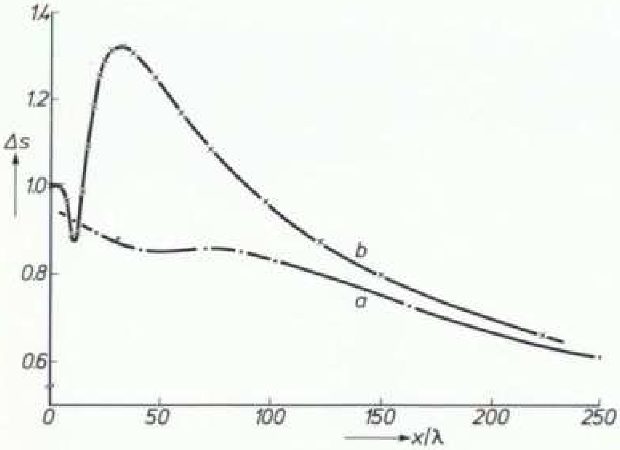


Fig. 14. De amplitudefouten Δs die door buiging ontstaan in het signaal zoals dat op verschillende afstanden van de zendende omzetter ontvangen wordt door een uit één element bestaande ontvanger. De zender heeft 14 elementen en een opening van 10λ. Kromme *a* heeft betrekking op een ontvanger met een opening van 10λ, kromme *b* op een ontvanger met een opening λ; x is de voortplantingsrichting, λ de golflengte. Men kan zien dat voor een kleine ontvangende elektrode het signaal in werkelijkheid in een zeker gebied door de buiging vergroot wordt. Wanneer ontvanger en zender dezelfde opening hebben, zijn de fouten betrekkelijk klein en daarom is het voordelig, de bronsterkteverdeling te bewerkstelligen door breedtevariaties.

De door ons uitgevoerde berekeningen hebben betrekking op isotrope oppervlakken, maar anderen [5] hebben aangetoond dat de resultaten veelal ook kunnen worden toegepast voor anisotrope materialen door eenvoudig de schaal van de voortplantingsweglengte te veranderen: het intensiteitsprofiel dat men op een afstand van x golflengten van de zender vindt in het isotrope geval, treft men in plaats daarvan aan op een afstand ax in een anisotroop geval, waarbij a groter of kleiner dan 1 kan zijn, afhankelijk van de anisotropie. Wij hebben dit proefondervindelijk geverifieerd.

Om voor de hand liggende redenen is de berekening van buigingseffecten op enkele golflengten afstand van een straler geen probleem dat gewoonlijk in verhandelingen over optische buiging wordt behandeld. Een formulering van het probleem in de termen van de functies van Green leidt tot de volgende vergelijking voor de buiging van oppervlaktegolven op een isotroop oppervlak:

$$\psi = -\frac{jkc}{2\pi\sqrt{r}}\left[H_1^{(2)}(kr) + \frac{1}{2kr}H_0^{(2)}(kr)\right]\mathrm{d}S.$$

Hierin is ψ de amplitude in een punt op een afstand r van een kleine lijnvormige bron $\mathrm{d}S$, $k = 2\pi/\lambda$, $H_0^{(2)}$ en $H_1^{(2)}$ zijn Hankelfuncties, en c is een constante. Wanneer r groter is dan ongeveer een golflengte, kan dit worden vereenvoudigd tot

$$\psi = \frac{c'k}{\sqrt{(kr)}}\exp \mathrm{j}(\omega t - kr + \pi/4).$$

Het door een zender met lengte $2l_1$ afgegeven signaal dat door een ontvanger met lengte $2l_2$ op d golflengten afstand wordt ontvangen is dan:

$$c'k\int_{-l_1}^{l_1}\int_{-l_2}^{l_2}\frac{d^2\exp \mathrm{j}(\omega t - kr + \pi/4)}{r^2\sqrt{(kr)}}\,\mathrm{d}x\,\mathrm{d}x',$$

waarin $r^2 = d^2 - (x - x')^2$. Uitdrukkingen van dit soort moeten voor elke elektrode in zowel de zender als de ontvanger worden uitgewerkt om het effect op het gehele filter te verkrijgen.

De keuze van het substraatmateriaal

Wanneer men een oppervlaktegolffilter ontwerpt, is het van groot belang een substraatmateriaal te kiezen waarvoor de koppeling tussen de akoestische oppervlaktegolf en het door de omzetters opgewekte elektrische veld zo groot mogelijk is. Zowel directe berekening als experimentele meting nemen veel te veel tijd in beslag, want men wenst immers een groot aantal mogelijke voortplantingsrichtingen in elk van een aantal materialen te onderzoeken. Gelukkig is er een handige benaderingsmethode voorgesteld [6] [7] die wij zeer bruikbaar hebben bevonden. Het idee is hierbij dat een oppervlaktegolf in een piëzo-elektrisch materiaal vergezeld gaat van een elektrisch veld waarvan de horizontale component is kortgesloten wanneer het oppervlak van het materiaal met een dunne geleidende laag is be-

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

dekt. Hierdoor wordt de snelheid van de golf gewijzigd, en deze wijziging is een bruikbare maat voor de wisselwerking tussen elektrische en akoestische velden in het oppervlaktegebied. De berekening is niet eenvoudig, maar is een goed aansluitende uitbreiding van een reeds bij MRL voorhanden rekenprogramma [8] voor het bepalen van de kenmerken van de voortplanting van oppervlaktegolven op piëzo-elektrische materialen met willekeurige anisotropie. *Fig. 15* geeft enige kenmerkende resultaten van onze berekeningen.

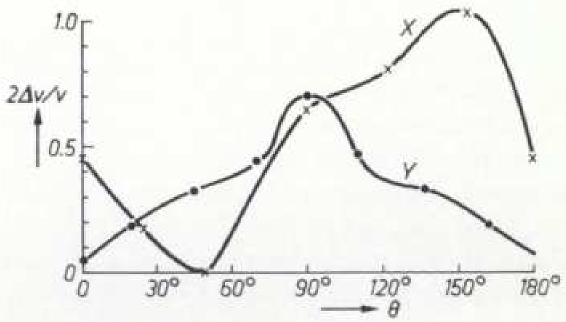


Fig. 15. Richtingsafhankelijkheid van de koppeling (berekend) tussen het elektrische veld en de akoestische oppervlaktegolf in een substraat van lithiumtantalaat. Een algemeen gebruikelijke maat voor de koppeling is $2\Delta v/v$, waarbij v de snelheid van de golf is en Δv de snelheidsverandering bij kortsluiting van het horizontale elektrische veld door een geleidende laag. Kromme X heeft betrekking op een substraat met X-snede, en voor deze kromme moet θ worden gelezen als de hoek tussen de voortplantingsrichting en de X-richting van het kristal. Kromme Y heeft betrekking op een substraat met Y-snede; hier is θ de hoek tussen de voortplantingsrichting en de Y-richting van het kristal.

Er moeten nog vele andere factoren in de beschouwing worden betrokken voordat de keuze van het substraatmateriaal en van de voortplantingsrichting kan worden bepaald. Wij noemen er hier enkele:

1. De diëlektrische constante. Deze bepaalt de impedantie van de omzetter; in het bijzonder voor grote elektrodenconfiguraties, is het van belang dat de diëlektrische constante laag is als men de impedantie redelijk hoog wil houden.
2. De toestand van het oppervlak moet fotolithografische behandeling toelaten.
3. De snelheid moet reproduceerbaar zijn en niet te veel veranderen met de tijd of de temperatuur. Snelheidsveranderingen zullen ook de centrale frequentie van het filter in dezelfde verhouding wijzigen. Voor een mf.-televisiefilter is een temperatuurcoëfficiënt van ongeveer $10^{-4}/°C$ voldoende goed, maar voor filters met smalle doorlaatband zijn de eisen veel strenger. De temperatuurcoëfficiënt verandert uit de aard der zaak gewoonlijk met de richting van voortplanting, en ook voor deze eigenschap kan de optimale richting van voortplanting het handigst door berekening worden bepaald.
4. In anisotrope materialen staat de richting van voortplanting niet noodzakelijkerwijs loodrecht op het golffront, zodat de oppervlaktegolf niet loodrecht op de elektrodes van een zendende omzetter wordt uitgezonden. Dit kan bij het ontwerp van een filter hinderlijk zijn, en het is gewoonlijk voordelig om een richting te kiezen waarvoor voortplanting en golfnormaal althans bij benadering dezelfde richting hebben. Ook dit kan alweer het handigst door berekening worden gedaan.

Middenfrequent-televisiefilter

Wij zullen nu in het kort de uitkomsten beschrijven, verkregen door toepassing van deze methodes op het probleem van de mf.-selectiviteit voor televisie. Dit is voor oppervlaktegolven in vele opzichten een ideale toepassing. Het vereiste filter moet een betrekkelijk grote bandbreedte hebben en de centrale frequentie heeft een zodanige waarde dat de elektrodenconfiguraties zonder moeilijkheid met behulp van gangbare fotolithografische methodes kunnen worden gemaakt. De doorlaatkromme is voldoende gecompliceerd om een volledig gebruik van de mogelijkheden tot bronsterkteverdeling nodig te maken, en het heeft grote voordelen het filter zo te maken dat het gemakkelijk in een capsule kan worden ondergebracht.

Het onderzoekprogramma heeft ten doel na te gaan of het mogelijk is een filter te maken dat goedkoop is en dat samen met zijn voorversterker en met zo veel mogelijk van de rest van de mf.-schakeling kan worden ondergebracht in een capsule voor een geïntegreerde schakeling. Zo'n capsule heeft het bijkomende voordeel dat er geen afregeling nodig is, noch in de fabriek, noch daarna. Zo'n filter zou een aantrekkelijk onderdeel zijn, mits aan de gestelde eisen wordt voldaan en de prijs laag is. Het tegemoetkomen aan de commerciële eis van een lage prijs, vooral waar het het onderbrengen in een capsule betreft, werpt een reeks problemen op die alle moeten worden opgelost voordat men tot een bruikbaar onderdeel kan komen. In vele opzichten vormen deze problemen een even grote uitdaging als de eigenlijke synthese van de doorlaatkromme.

Monokristallijne piëzo-elektrische materialen zijn ideaal als substraten omdat ze zeer gelijkmatig en reproduceerbaar zijn, en betrekkelijk vrij van verouderingsverschijnselen. Prijsoverwegingen leggen echter groot gewicht in de schaal, al doen zij die niet altijd doorslaan,

[5] I. M. Mason, Two-dimensional surface-wave diffraction from an arbitrary source in an anisotropic medium, Electronics Letters 7, 344-345, 1971 (no. 12).
[6] J. J. Campbell en W. R. Jones, A method for estimating optimal crystal cuts and propagation directions for excitation of piezoelectric surface waves, IEEE Trans. **SU-15**, 209-217, 1968.
[7] K. A. Ingebrigtsen, Surface waves in piezoelectrics, J. appl. Phys. **40**, 2681-2686, 1969.
[8] R. G. Pratt, G. Simpson en W. A. Crossley, Acoustic-surface-wave properties of $Bi_{12}GeO_{20}$, Electronics Letters **8**, 127-128, 1972 (no. 5).

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

ten gunste van synthetische piëzo-elektrische keramische materialen, de goedkoopste die aan de gestelde eisen voldoen. Helaas is de piëzo-elektrische keramiek meestal enigszins poreus en dit bemoeilijkt het verkrijgen van een voldoend nauwkeurig gedefinieerd elektrodenpatroon. Bovendien zijn ze niet altijd reproduceerbaar. Door werk van Philips Forschungslaboratorium Aachen, de Hoofdindustriegroep Elcoma en Philips' Natuurkundig Laboratorium is het probleem van de porositeit echter opgelost, zodat wij nu over een keramisch materiaal beschikken dat vrijwel geheel aan de gestelde eisen voldoet. De diëlektrische constanten van keramische materialen zijn echter hoog, en om de filterimpedantie op een redelijke waarde te houden hebben we dan ook met akoestische openingen moeten werken van niet meer dan 10 golflengten. Daarbij is de capacitantie van het filter echter nog voldoende laag voor de geïntegreerde uitvoering van de speciale voorversterker, en om het vereiste niveau (15 pF) te bereiken moeten twee kleine filters in serie worden geschakeld.

Parasitaire signalen

In een oppervlaktegolffilter treden vele soorten van parasitaire signalen op: daartoe behoren reflecties van oppervlaktegolven aan de randen van het substraat, resonantietrillingen van het substraat in zijn geheel (in een andere modus dan die der oppervlaktegolven) en signalen afkomstig van ruimtelijke akoestische golven

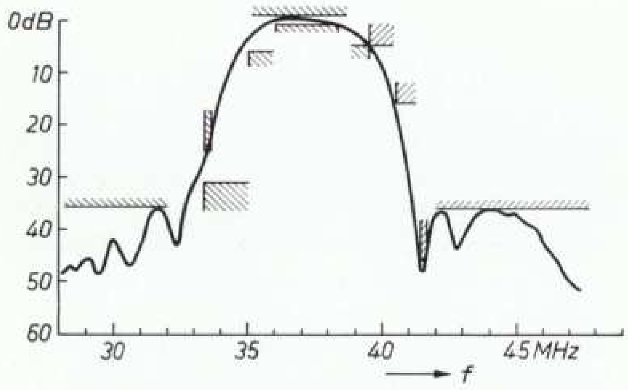


Fig. 16. Doorlaatkromme van een mf.-televisiefilter met oppervlaktegolven, vergeleken met de specificatie. Dit filter is, met zijn eigen voorversterker, gemonteerd in een capsule voor een geïntegreerde schakeling. De gearceerde vlakken geven de in de specificatie vastgelegde randwaarden aan. Bij het vaststellen van deze randwaarden is rekening gehouden met de selectiviteit van de mf.-spoel die normaliter deel uitmaakt van een televisieontvanger.

die eveneens worden uitgezonden door de omzetterconfiguratie. Met al deze parasitaire signalen moet rekening worden gehouden en ze moeten worden teruggebracht tot een niveau waarop ze geen nadelige invloed uitoefenen.

Rechtstreekse elektrische overdracht der signalen vormt een probleem op zichzelf. Het filter heeft een tussenschakeldemping van ongeveer 23 dB en bij het diepste minimum moet de demping 52 dB bedragen. Het niveau van het rechtstreeks elektrisch overgebrachte signaal moet dus 75 dB beneden het ingangssignaal liggen, een eis waaraan binnen een zo klein bestek niet gemakkelijk is te voldoen.

Al deze problemen worden nog verzwaard door de noodzaak, het substraat zo klein mogelijk te houden, om daardoor de prijs te drukken en het onderbrengen in een capsule te vergemakkelijken. Desondanks hebben wij reeds filters gemaakt die de gestelde eisen zeer dicht benaderen, al zijn er nog zorgen over sommige van de parasitaire ruimtegolven. Deze filters worden gemonteerd in een capsule voor een geïntegreerde schakeling, samen met een speciaal ontworpen voorversterker waarmee zij een complete filtereenheid vormen. In *fig. 16*, de doorlaatkromme van zo'n filter, zijn tevens de in de specificatie vastgelegde grenzen aangegeven die, er zij op gewezen, rekening houden met de selectiviteit van de mf.-spoel die men normaliter in een televisieontvanger aantreft. De tussenschakeldemping van het filter is 23 dB, zijn capacitantie 15 pF en zijn afmetingen zijn $8 \times 3 \times 0{,}5$ mm. Dit speciale exemplaar is gemaakt op een substraat van kaliumnatriumniobaat, een keramisch materiaal waarvoor de voortplantingssnelheid een te hoge temperatuurcoëfficiënt heeft. Het nieuwe keramische materiaal heeft de vereiste temperatuurcoëfficiënt; wij zijn thans filters op dit materiaal aan het maken, evenals op twee monokristallijne materialen die geschikte alternatieven zouden kunnen vormen.

De schrijver zou er nadruk op willen leggen dat veel van het in dit artikel beschreven werk gedaan is door anderen in het team waarmee hij bij MRL is verbonden. In deze zin is hij ook de vertegenwoordiger van: M. Chapman, Eileen Hughes, K. W. Moulding, J. S. Palfreeman, D. W. Parker, R. G. Pratt, G. Simpson, J. S. Singleton en R. Stevens.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Optische polarisatie-effecten in een gaslaser

H. de Lang, D. Polder en W. van Haeringen

Elk boek over klassieke optica bevat wel een belangrijk hoofdstuk over de karakteristieke eigenschappen van gepolariseerd licht en zijn voortplanting in „fysische" media. Experimenten met gepolariseerd licht hebben ook veel bijgedragen tot het begrijpen van de structuur van atomen, moleculen en materie. De komst van de laser heeft een herleving van de optica veroorzaakt, vooral omdat deze ons de beschikking gaf over lichtbronnen waarvan intensiteit en coherentie in ruimte en tijd bijzonder groot zijn. Dit artikel houdt zich bezig met de interactie van laserlicht met het medium waarin het gegenereerd wordt, en in het bijzonder met de optische polarisatie-effecten welke veroorzaakt worden door deze interactie.

Het eerste van de drie volgende hoofdstukken bevat een algemene uiteenzetting van de belangrijkste begrippen uit de fysica van lasers, die voor het begrijpen van de rest van het artikel nuttig kan zijn. In het volgende hoofdstuk wordt de informatie besproken welke uit de experimentele studie van polarisatie-effecten in een gaslaser kan worden verkregen. Het onthult onorthodox optisch gedrag van het genererend medium en bevat een fenomenologische en klassieke beschrijving daarvan. In het laatste hoofdstuk worden de basiselementen voor een kwantitatieve beschrijving gegeven. Het geeft aan hoe een quantummechanische theorie de waarnemingen tot in details doet begrijpen en leidt tot enkele voorspellingen die bevestigd worden door het experiment.

Algemene begrippen

Lasers en masers zijn apparaten waarin gebruik gemaakt wordt van de versterking die elektromagnetische golven ondervinden in daartoe geschikt gemaakte gasvormige, vloeibare of vaste media. De afneming in amplitude van een zich in een medium voortplantende elektromagnetische golf is een welbekend verschijnsel: de absorptiecoëfficiënt α van het medium is gedefinieerd als de relatieve afneming per eenheid van lengte $-I^{-1}\,\mathrm{d}I/\mathrm{d}x$ van de intensiteit I van de golf. Een toeneming in amplitude, d.w.z. versterking, kan worden verkregen door aan de absorptiecoëfficiënt van het medium een negatieve waarde toe te kennen. Het is mogelijk om media te maken die deze ongebruikelijke eigenschappen bezitten. De negatieve waarde van α wordt dan veroorzaakt door een verschijnsel dat later besproken zal worden: gestimuleerde energie van straling. Het mechanisme dat de negatieve absorptie doet ontstaan, wordt het pompmechanisme genoemd. De energie die nodig is om een golf te versterken wordt door het medium aan de golf geleverd, en daarom moet er een externe energiepomp aanwezig zijn die energie aan het medium toevoert.

In tegenstelling tot hetgeen de woorden LASER (Light Amplification by Stimulated Emission of Radiation) of MASER (Microwave Amplification by SER) zouden kunnen suggereren, kan in principe elke soort elektromagnetische golf met daartoe geschikte inrichtingen door middel van gestimuleerde emissie worden versterkt. In de praktijk is bij vele frequenties tussen die van laagfrequente microgolven en kortgolvig ultraviolet licht versterking gerealiseerd. De verschillende daarvoor gebruikte media hebben zeer specifieke eigenschappen, en in het algemeen gesproken zal een bepaald medium met bijbehorend pompmechanisme alleen in een klein frequentiegebied elektromagnetische golven versterken. Dit beschikbare spectrale gebied kan van geval tot geval sterk verschillen.

Eenmaal in de toestand van negatieve absorptie gebracht, kan het medium gebruikt worden om de corresponderende elektromagnetische golven te genereren. Dit generatieproces begint bij de in het medium steeds aanwezige ruis die door de versterking tot een grote intensiteit kan aangroeien. Daarbij kan dan het terugkoppelingsprincipe worden toegepast, door de groeiende golf een aantal malen opnieuw het medium te laten passeren. De straling die op die manier gegenereerd wordt, heeft een frequentie die ergens in het beschikbare spectrale gebied ligt en kan buitengewoon goed monochromatisch zijn, met een spectrale zuiverheid die vele ordes van grootte beter is dan ooit op enigerlei andere manier bereikt is.

In het optische gebied, van het verre infrarood tot het ultraviolet, bieden verscheidene door een gasontlading „gepompte" gassen een grote verscheidenheid van mogelijke laserfrequenties. Het beschikbare spectrale gebied rond elke frequentie is meestal erg klein. In de helium-neon-laser bij voorbeeld, die bij verscheidene frequen-

Prof. D. Polder, wetenschappelijk adviseur van Philips' Natuurkundig Laboratorium, Eindhoven; dr. H. de Lang en dr. W. van Haeringen, medewerkers van hetzelfde laboratorium.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

ties in het rood en infrarood kan werken, is bij centrumfrequenties van ongeveer 3×10^{14} Hz het beschikbare gebied van de orde van 500 MHz. Er zijn vele vaste stoffen die, als ze gepompt worden door een gepulste lichtbron van grote intensiteit, laserwerking vertonen in het optische gebied. Voorbeelden zijn het robijnkristal en glas waarin zich geschikte ionen bevinden, zoals neodymium. Lasers worden meestal gebruikt als generators van licht met een grote helderheid en een zeer kleine lijnbreedte. Gaslasers zijn unieke lichtbronnen met een buitengewone coherentie in de ruimte en de tijd. In de holografie wordt van deze eigenschappen gebruik gemaakt. Vastestoflasers kunnen licht met een uiterst hoog vermogen genereren, vooral als de energie wordt vrijgemaakt in pulsen van slechts 10^{-8} tot 10^{-12} seconde. Gedurende een korte puls kunnen vermogens tot 10^9 watt worden bereikt. Geheel nieuwe effecten treden aan de dag als materie wordt bestraald met zo'n hoog vermogen.

We zullen hier niet trachten een samenvatting van alle verschillende laserbronnen te geven, maar gaan nu over tot een discussie van het verschijnsel der gestimuleerde emissie.

Gestimuleerde emissie

In de quantumtheorie wordt de absorptie van elektromagnetische golven verklaard door de overgang van een atoom of molecuul van een toestand (a) van lagere energie naar een toestand (b) van hogere energie onder de invloed van invallend licht. Als het verschil in energie tussen de twee toestanden ΔE is, dan kan alleen licht worden geabsorbeerd met een frequentie $\nu_{ba} = \Delta E/h$, waarin h de constante van Planck is. De waarschijnlijkheid dat een atoom dat aanvankelijk in de toestand (a) verkeerde, naar (b) overgaat, is evenredig met de intensiteit I van het licht. De waarschijnlijkheid dat een atoom aanvankelijk in de toestand (b) onder de invloed van licht overgaat naar de lagere toestand (a), is eveneens evenredig met I en is in feite, als alle andere factoren hetzelfde zijn, gelijk aan de eerder genoemde waarschijnlijkheid. Wanneer onder invloed van invallend licht een overgang van een bovenniveau naar een benedenniveau plaatsvindt, wordt de vrijkomende energie toegevoegd aan die van het invallende licht. Dit is de gestimuleerde emissie van straling. Als er per eenheid van volume N_a atomen in de toestand (a) en N_b in (b) zijn, dan zal de afneming $-I^{-1}\,dI/dx$ van de intensiteit evenredig zijn met $N_a - N_b$, hetgeen betekent dat er alleen versterking kan optreden als $N_b > N_a$ is.

Dit kan nauwkeuriger worden geformuleerd. Per volume-element induceert het elektrische veld $E = E_0 \cos \omega t$ van het licht een diëlektrische polarisatie P per eenheid van volume die wordt gegeven door

$$P = P_{in} \cos \omega t + P_{out} \sin \omega t \qquad (1)$$

en die dezelfde hoekfrequentie $\omega = 2\pi\nu$ heeft als het licht. Zowel de component P_{in}, in fase met E, als de uit-fase-component P_{out} is evenredig met E_0 en ook met $N_a - N_b$. Het deel P_{out} dat $\pi/2$ in fase achter is, is positief als $N_a > N_b$ is, hetgeen betekent dat het medium energie absorbeert, terwijl P_{out} negatief is als $N_a < N_b$ is, zodat in dat geval de lichtgolf wordt versterkt. Zowel P_{in} als P_{out} hangt van ω af. In het bijzonder is $|P_{out}|$ een klokvormige functie van ω met een maximum in $\omega = \omega_{ba}$ ($= 2\pi\nu_{ba}$), hetgeen samenhangt met het feit dat een atomaire absorptielijn om diverse redenen een niet-verwaarloosbare spectrale breedte heeft. Als de absorptie negatief is, is binnen deze breedte versterking en laserwerking mogelijk. In het bovenstaande werd deze breedte het beschikbare spectrale gebied genoemd. De waarde en de frequentieafhankelijkheid van P_{in} bepalen respectievelijk de brekingsindex en de dispersie van het medium, welke, zoals men weet, in de buurt van een atomaire overgang aanzienlijk kunnen zijn.

Tot nu toe is verondersteld dat N_a en N_b gegeven grootheden zijn. In een medium in thermodynamisch evenwicht zijn de atomaire niveaus van hogere energie altijd minder sterk bezet dan die van lagere energie, d.w.z. $N_a > N_b$. In een laser of maser moet een zeer specifiek pompmechanisme werkzaam zijn om voor ten minste één paar niveaus (a) en (b) een geïnverteerde bezetting ($N_a < N_b$) te verkrijgen. Pompmechanismen zullen we hier verder niet bespreken, omdat elk verschillend type laser of maser op zijn eigen wijze wordt gepompt. We zullen eenvoudig de aanwezigheid van een pomp aannemen die per eenheid van volume en per eenheid van tijd Λ atomen in de toestand (b) brengt. Zelfs bij afwezigheid van invallende straling zijn er vele processen, zoals botsingen tussen atomen, waardoor de atomen uit de toestand (b) naar bij voorbeeld geheel verschillende lager gelegen toestanden kunnen vallen. Uiteraard bemoeilijken zulke processen het totstandkomen van een geïnverteerde bezetting. We zullen veronderstellen dat het aantal atomen dat per eenheid van tijd op die manier uit (b) verdwijnt, evenredig is met het aantal dat aanwezig is in (b), d.w.z. gelijk is aan $\gamma_b N_b$. Zo veronderstellen we ook dat het aantal dat door overgangen naar lager gelegen energieniveaus per tijdseenheid uit (a) verdwijnt, gelijk is aan $\gamma_a N_a$. Deze laatste overgangen bevorderen een inversie van de bezetting tussen (a) en (b) maar zijn natuurlijk alleen effectief als (a) niet de laagste toestand van het atoom is. Ter vereenvoudiging zullen we aannemen dat γ_a zo groot is dat N_a onder alle omstandigheden verwaarloosbaar blijft ($N_a \approx 0$). Ten slotte zijn er overgangen van (b) naar (a) die de energie leveren welke voor de ver-

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

sterking van de invallende lichtgolf vereist is. De overgangswaarschijnlijkheid voor deze gestimuleerde emissie wordt gegeven door $c_1(N_b - N_a)I$, en is dus evenredig met $N_b - N_a$ en de lichtintensiteit. Als $N_a \approx 0$ is, voldoet N_b aan de volgende differentiaalvergelijking:

$$\mathrm{d}N_b/\mathrm{d}t = \Lambda - \gamma_b N_b - c_1 N_b I. \qquad (2)$$

Voor gegeven I volgt hieruit de stationaire over bezetting:

$$N_b - N_a \approx N_b = \Lambda/(\gamma_b + c_1 I). \qquad (3)$$

De grootheid c_1 in deze vergelijkingen is een functie van de lichtfrequentie ν.

Uit deze elementaire beschouwing kunnen twee conclusies worden getrokken. In de eerste plaats is de negatieve absorptiecoëfficiënt, die evenredig was met $N_b - N_a$, als gevolg van (3) een functie van I. Zijn absolute waarde neemt af met toenemende I, een effect dat *verzadiging van het medium* genoemd wordt. Dit betekent dat een elektromagnetische golf die men steeds maar weer opnieuw, bij voorbeeld met behulp van spiegels, het medium laat passeren, niet steeds blijft doorgroeien. De afneming van de versterking leidt tot een stabilisatie van de intensiteit en de daarmee corresponderende negatieve absorptiecoëfficiënt, en wel op een waarde waarvoor de versterking van de golf juist gecompenseerd wordt door de altijd aanwezige verzwakking ten gevolge van effecten als absorptie in of transmissie door de spiegels. Dan hebben we een laser die als een door zijn intensiteit gestabiliseerde oscillator werkt. In de tweede plaats wordt de in het medium geïnduceerde diëlektrische polarisatie P, die evenredig met E_0 en $N_b - N_a$ was, een niet-lineaire functie van E_0, omdat I in (3) zelf evenredig is met E_0^2. Voor kleine E_0, d.w.z. als $E_0^2 \sim I \ll \gamma_b/c_1$, zal een reeksontwikkeling van P in opklimmende oneven machten van E_0 de niet-lineariteit benaderen. Later in dit artikel zal de niet-lineariteit een belangrijke rol spelen in situaties waarin het vectorkarakter van de grootheden E en P in aanmerking genomen moet worden.

In het voorgaande hebben we de spontane emissie niet besproken. Het mogelijk zijn van absorptie en gestimuleerde emissie door middel van overgangen tussen toestanden (a) en (b) impliceert, ongeacht de aanwezigheid van invallend licht, het bestaan van spontane emissie van licht met frequentie ν_{ba} door een spontane overgang van (b) naar (a) steeds als het atoom in de toestand (b) is. Aangezien de overgangswaarschijnlijkheid voor spontane emissie echter evenredig is met N_b, kan het effect daarvan op vergelijking (2) verdisconteerd worden door aanpassing van de keuze van γ_b. Bovendien draagt spontaan geëmitteerd licht niet bij tot de versterking van de invallende golf. Dit licht is daartoe niet in staat door de richting waarin het wordt uitgezonden of door gebrek aan coherentie met het invallende licht.

Spontane emissie is in een stationair oscillerende laser een bron van ruis. Zulke effecten zullen we in dit artikel niet bespreken.

Karakteristieke eigenschappen van de gaslaser

Omdat de effecten die in een continu werkende gaslaser optreden in dit artikel later besproken zullen worden, zullen we nu een overzicht geven van de belangrijkste eigenschappen van zo'n laser. Beschouw een helium-neon-laser die oscilleert op een golflengte $\lambda_l = c/\nu_l$ ($\lambda \approx 1{,}15\ \mu$m, c is de lichtsnelheid) en die bestaat uit twee onderling evenwijdige vlakke ronde spiegels met een diameter d van enkele mm en een onderlinge afstand L van ongeveer 10 cm met daartussen een cilindrische buis welke een helium-neon-mengsel bevat (90% He; 10% Ne; druk 5 torr, d.w.z. ca. 10^{16} Ne-atomen/cm^3). De spiegels hebben een reflectiecoëfficiënt van ca. 0,99. Het gasmengsel wordt „gepompt" door middel van een elektrische gasontlading, waardoor botsingen tussen helium- en neonatomen een geïnverteerde bezetting opbouwen tussen een paar energieniveaus (b) en (a) van de neonatomen, met een energieverschil $\Delta E = h\nu_{ba} \approx hc/\lambda_l$. Een karakteristieke waarde voor $N_b - N_a$ is 10^9/cm^3, d.w.z. slechts een zeer kleine fractie van het totale aantal atomen. De frequentieafhankelijkheid van de resulterende onverzadigde negatieve absorptiecoëfficiënt wordt gegeven door:

$$\alpha(\omega) = \alpha(\omega_{ba}) \exp\left[-(\omega - \omega_{ba})^2/\gamma_D^2\right], \qquad (4)$$

met $\alpha(\omega_{ba}) < 0$ en $|\alpha(\omega_{ba})| \approx 10^{-3}$ cm^{-1}. De spectrale breedte van α, waarvoor $\gamma_D \approx 10^9$ s^{-1} een maat is, wordt veroorzaakt door het Doppler-effect, dat later besproken zal worden.

De relatieve winst in intensiteit van een tussen op afstand L geplaatste spiegels heen en weer gaande lichtgolf met hoekfrequentie ω_l is bij één doorgang gelijk aan $|\alpha(\omega_l)| \times L$ en bedraagt ongeveer 1%. Anderzijds is er een relatief verlies $\mathcal{L}$ per cyclus, hetgeen veroorzaakt wordt door transmissie en absorptie bij de spiegels en doordat een zeer klein deel van de lichtbundel verloren gaat langs de randen van de spiegel. Zolang $2L \times |\alpha(\omega_l)| > \mathcal{L}$ is, wordt de bundel netto versterkt. In de stationaire situatie is de negatieve absorptiecoëfficiënt bij $\omega = \omega_l$ gereduceerd tot een waarde α_s, die voldoet aan:

$$2\,|\alpha_s(\omega_l)| = \mathcal{L}/L. \qquad (5)$$

De verhouding van onverzadigde tot verzadigde waarde wordt gegeven door:

$$F = \alpha(\omega_l)/\alpha_s(\omega_l), \qquad (6)$$

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

en $F-1$ wordt de overmaatfractie boven de drempelwaarde genoemd. Hoe groter $F-1$ is, des te groter wordt de intensiteit I van de staande lichtgolf in de stationaire oscillator.

Zodra $2|\alpha(\omega_{\rm ba})| > \mathcal{L}/L$ wordt, bestaat er een frequentiegebied rond $\omega_{\rm ba}$ waarvoor geldt:

$$2|\alpha(\omega)| > \mathcal{L}/L. \qquad (7)$$

De precieze frequentie $\omega_l = 2\pi\nu_l$ waarop de laser in feite zal oscilleren, wordt nu bepaald door de eis dat de lichtgolf na elke cyclus ter lengte $2L$ op constructieve wijze met zichzelf interfereert. Als de fasesprongen bij de reflecterende spiegels verwaarloosd worden, dan moet gelden:

$$2L = p\lambda_l = pc/\nu_l, \qquad (8)$$

waarbij het getal p een geheel getal is van de orde 2×10^5. In het algemeen bestaan er nu een zeker aantal frequenties ν_l, onderling gescheiden door frequentie-intervallen ter grootte $\Delta\nu_l = c/2L$,waarvoor zowel aan (7) als aan (8) voldaan is. Als $L \approx 10$ cm, dan is $\Delta\nu_l$ bij benadering gelijk aan $1{,}5\times10^9$ Hz, d.w.z. van dezelfde orde van grootte als de spectrale (Doppler-) breedte van α. Het is daarom mogelijk de laser te laten werken op één enkele frequentie ν_l, en ook om hem binnen het Doppler-profiel af te stemmen door een precieze aanpassing van L.

In het voorgaande werd min of meer stilzwijgend aangenomen dat de lichtgolf een vlakke golf is die tussen de spiegels heen en weer loopt. Dit kan niet geheel juist zijn, aangezien de bundel en de spiegels een eindige diameter d hebben (enige millimeters). In werkelijkheid zijn, aangezien $\lambda/d \ll 1$ en ook $d^2/\lambda L > 1$ is, de afwijkingen van een vlakke golf klein, maar het vereist een nauwkeuriger analyse om aan te tonen dat de vereenvoudiging tot vlakke golven voor ons doel is toegestaan. In zo'n analyse wordt het systeem van twee kleine spiegels op afstand L, die vlak of licht gekromd mogen zijn, beschouwd als een resonator voor elektromagnetische golven. Net als bij de meer bekende gesloten resonatoren voor microgolven, hebben open resonatoren een groot aantal, hier meestal dicht bij elkaar liggende, eigenfrequenties, elk met een zekere verzwakking. Indien er gepompt wordt ontdempt het medium een of meer van de corresponderende trillingswijzen met frequentie ν_l welke aan (7) en aan een meer verfijnde versie van (8) voldoen.

Men kan het optisch systeem ook opvatten als een speciaal type interferometer van Fabry-Perot. Door de (weliswaar geringe) doorlating van de spiegels vertoont de lege resonator voor een externe, evenwijdig met de optische as invallende lichtbundel als functie van de frequentie de voor een Fabry-Perot-interferometer karakteristieke doorlatingsmaxima. Voor ons doel is het voldoende om aan te nemen dat er in een laser die oscilleert op frequentie ν_l, een welgedefinieerd staandegolfpatroon bestaat, en dat dit patroon samengesteld gedacht kan worden uit twee tussen de spiegels voortlopende vlakke golven. Door de doorlating van de spiegels kan een smalle evenwijdige bundel het systeem verlaten, waaraan dan de experimentele waarnemingen kunnen worden gedaan.

Doppler-verbreding

Voor het juiste begrip van de werking van gaslasers is een discussie van de Doppler-verbreding $\gamma_{\rm D}$ in vergelijking (4) noodzakelijk. De Doppler-verbreding $\gamma_{\rm D}$ van de positieve of negatieve absorptiecoëfficiënt $\alpha(\omega)$, die een zich voortplantende golf ondervindt, vindt zijn oorsprong in de snelheidsverdeling van de atomen. Als alle atomen in de toestanden (a) en (b) in rust zouden zijn, dan zou een veel smallere, om $\omega_{\rm ba}$ geconcentreerde lijn $\alpha_0(\omega)$ worden verkregen:

$$\alpha_0(\omega) = A(N_{\rm a}-N_{\rm b})/[(\omega-\omega_{\rm ba})^2+\gamma_{\rm ba}{}^2]. \qquad (9)$$

Hierin is de karakteristieke breedte $\gamma_{\rm ba}$ veel kleiner dan $\gamma_{\rm D}$, en bovendien verschilt de vorm van de lijn van die volgens vergelijking (4). In (9) is $A > 0$ en is de evenredigheid van α met $N_{\rm a}-N_{\rm b}$ expliciet gemaakt.

Als alle atomen een snelheid v_z zouden hebben, dan zou een golf die zich voortplant in de $+z$-richting, absorptie ondervinden in een gebied rond een verschoven centrale frequentie ω_{v_z}, en wel volgens:

$$\alpha_{v_z}{}^+(\omega) = A(N_{\rm a}-N_{\rm b})/[(\omega-\omega_{v_z})^2+\gamma_{\rm ba}{}^2], \qquad (10)$$

met

$$\omega_{v_z} = \omega_{\rm ba}\,(1+v_z/c). \qquad (11)$$

Dit is een direct gevolg van het Doppler-effect zoals dat optreedt voor een waarnemer die zich met het atoom mee beweegt. In een gas bestaat een verdeling van atomaire snelheden v_z. Als $n_{\rm a}(v_z){\rm d}v_z$ het aantal atomen per eenheid van volume in de toestand (a) in het snelheidsinterval tussen v_z en $v_z+{\rm d}v_z$ is, en $n_{\rm b}(v_z){\rm d}v_z$ dat voor atomen in (b), dan geldt derhalve:

$$\alpha^+(\omega) = A\int {\rm d}v_z\,[n_{\rm a}(v_z)-n_{\rm b}(v_z)]/[(\omega-\omega_{v_z})^2+\gamma_{\rm ab}{}^2]. \qquad (12)$$

De speciale frequentieafhankelijkheid van de onverzadigde α, zoals uitgedrukt door (4), volgt nu uit (12) als, bij afwezigheid van verzadiging, de snelheidsverdelingsfuncties $n_{\rm a}(v_z)$ en $n_{\rm b}(v_z)$ verdelingen volgens Maxwell voor atomen met massa M en temperatuur T zijn. In dit geval is $\gamma_{\rm D}{}^2$ gelijk aan $2kT\omega_{\rm ba}{}^2/Mc^2$, waarin k de constante van Boltzmann is.

Het Doppler-profiel $\alpha(\omega)$ is het resultaat van een gewogen superpositie van vele smalle lijnen met breedte $\gamma_{\rm ba}$. Deze breedte wordt de *homogene lijnbreedte* ge-

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

noemd omdat hij, in tegenstelling met de inhomogene Doppler-verbreding, niet verder kan worden ontbonden in bijdragen die afkomstig zijn van verschillende groepen atomen. Een van de oorzaken van de homogene verbreding is de als gevolg van botsingen en spontane emissie eindige levensduur van de niveaus. In de helium-neon-lasers is $\gamma_{ba} \approx 2 \times 10^7\ s^{-1}$, dat is inderdaad kleiner dan γ_D.

Ook de diëlektrische polarisatie P, die door het elektrische veld E van een golf in een gasvormig medium wordt geïnduceerd, bestaat uit de som van bijdragen van atomen in de verschillende snelheidsintervallen. Dit nu maakt de berekening van verzadigingseffecten, zoals de verzadiging van de negatieve absorptie of, meer algemeen, de berekening van de niet-lineaire relatie tussen P en E, aanzienlijk ingewikkelder. Hiervoor is een gedetailleerde kennis vereist van de wijze waarop een bundel van grote intensiteit en met frequentie ν_l voor elke snelheid de waarden van $n_a(v_z)$ en $n_b(v_z)$ beïnvloedt. Het punt is hier namelijk dat de bundel die in de positieve z-richting loopt, alléén de geïnverteerde bezetting $n_b - n_a$ reduceert van die atomen waarvoor ω_{v_z} in vergelijking (11) bij benadering gelijk is aan ω_l, d.w.z.:

$$|\omega_{v_z} - \omega_l| \leqslant \gamma_{ab}. \qquad (13)$$

Volgens (11) betekent dit dat alleen atomen waarvoor

$$v_z \approx c(\omega_l - \omega_{ba})/\omega_{ba} \qquad (14)$$

is, door de verzadiging worden beïnvloed. Om dezelfde reden beïnvloedt de bundel die in de negatieve z-richting terugloopt alleen atomen waarvoor

$$v_z \approx -c(\omega_l - \omega_{ba})/\omega_{ba}. \qquad (15)$$

Daardoor worden er, in een stationair oscillerende laser, twee „gaten" in de geïnverteerde verdeling $n_b(v_z) - n_a(v_z)$ „gebrand" (het hole-burning effect).

Een kwantitatieve beschrijving hiervan kan worden verkregen door voor de atomen in elk snelheidsinterval afzonderlijk, vergelijkingen van het type (2) en (3) op te stellen. Door het hole-burning effect wordt $|\alpha(\omega_l)|$ gereduceerd tot zijn verzadigingswaarde $|\alpha_s(\omega_l)|$; bovendien beïnvloeden beide „gaten" in de verdeling de brekingsindex voor elk van de twee golven. Als $\omega_l \approx \omega_{ba}$ is, treden zeer bijzondere effecten op, omdat dan atomen waarvoor $v_z = 0$ is, gelijktijdig door de twee golven worden verzadigd, zodat de laser-intensiteit dan lager zal zijn dan anders. Dit verschijnsel staat bekend als de „Lamb dip" in de intensiteit van de laserstraling bij $\omega_l \approx \omega_{ba}$.

In het volgende zal van de algemene begrippen die in het voorgaande belicht zijn, gebruik worden gemaakt, maar de aandacht zal gericht worden op de vragen die betrekking hebben op het vectorkarakter van het elektrische veld in een lichtgolf. Een stationair oscillerende laser produceert gewoonlijk slechts licht van één bepaalde optische polarisatietoestand. Als het lasersysteem cilindersymmetrisch is om de z-as, verwacht men niet dat het laserlicht in een bepaalde richting, b.v. de x-richting, optisch gepolariseerd zal zijn. Polarisatie in de y- of willekeurig welke andere richting zou even waarschijnlijk zijn, evenals trouwens elke superpositie van lineaire polarisaties, d.w.z. elliptische polarisatie, dat zou zijn. Zowel experimenteel als theoretisch blijkt dat deze overwegingen niet toereikend zijn en dat de experimenteel waargenomen polarisatieverschijnselen niet uitsluitend het gevolg zijn van afwijkingen van cilindersymmetrie, maar ook verband houden met de niet-lineaire eigenschappen die het medium heeft onder de invloed van gepolariseerd licht.

Optische polarisatietoestanden in laser-interferometers

Zoals boven uiteengezet, zou er geen enkele beperkende voorwaarde zijn met betrekking tot de polarisatietoestand van een trillingswijze in een isotrope, d.w.z. cilindersymmetrische laser. In de praktijk bestaat een dergelijke ideale situatie niet. Zelfs als de interferometer alleen bestaat uit twee vlakke evenwijdige spiegels zal er altijd wel een kleine anisotropie zijn, die bij voorbeeld veroorzaakt kan worden door onvolkomenheden in de reflecterende lagen op de spiegels. Inderdaad werd in een artikel [1] over de eerste helium-neon-laser, die van het Fabry-Perot-type was, gemeld dat de straling vlak gepolariseerd was in een richting die ogenschijnlijk bepaald werd door een lineair patroon van barstjes in die lagen. Toentertijd werd algemeen aangenomen dat de polarisatierichting de richting was waarvoor de interferometer een minimaal verlies vertoonde. Een dergelijke interpretatie in termen van absorptie-anisotropie impliceert dat de polarisatierichting onafhankelijk van de afstemcondities zou moeten zijn.

Uit experimenten met een kleine Fabry-Perot-helium-neon-laser ($\lambda = 1{,}15\ \mu m$) [2] in Philips' Natuurkundig Laboratorium bleek echter dat als de laser wordt verstemd, de polarisatierichting min of meer abrupt met $\pi/2$ verandert [3] [4] als men het centrum van het Doppler-profiel passeert. Deze polarisatiesprong, welke ook door andere onderzoekers werd gemeld, heeft ons genoodzaakt de interpretatie in termen van absorptie-anisotropie te verwerpen. We hebben haar vervangen door een interpretatie waarin zowel de

[1] D. R. Herriott, J. Opt. Soc. Amer. **52**, 31, 1962.
[2] H. G. van Bueren, J. Haisma en H. de Lang, Physics Letters **2**, 340, 1962. Zie ook J. Haisma, S. J. van Hoppe, H. de Lang en J. van der Wal, Philips techn. T. **24**, 217, 1962.
[3] H. de Lang, Physica **33**, 163, 1967 (Proc. Zeeman Centennial Conference, Amsterdam 1965).
[4] H. de Lang en G. Bouwhuis, Physics Letters **19**, 481, 1965.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

fase-anisotropie, d.w.z. het bestaan van verschillende optische weglengten voor verschillende polarisatietoestanden, als de door verzadigingseffecten geïnduceerde anisotropie van het gasvormige medium een rol speelt [3] [4] [5]. In het volgende zal worden uiteengezet hoe experimentele resultaten ons leidden tot het invoeren van het begrip verzadigingsanisotropie en hoe dit kan worden uitgedrukt in termen van de atomaire eigenschappen van het medium.

Om in te zien hoe de interferometer-anisotropie de polarisatietoestand van een trillingswijze beïnvloedt, maken we weer gebruik van de herhalingseigenschappen van de vrijwel vlakke golf, die tussen de spiegels heen en weer loopt. In een stationaire trillingswijze moet de golf, na een cyclus, a) zijn fase, b) zijn amplitude en c) zijn polarisatietoestand reproduceren. Als het ordegetal p van de golf gegeven is, legt conditie (a) de frequentie van de golf vast. Aan conditie (b) kan alleen worden voldaan wanneer er een versterkend medium aanwezig is. Conditie (c) zal opnieuw geformuleerd worden door gebruik te maken van de complexe amplitudes van de lineair gepolariseerde componenten E_x en E_y, waarin de elektrische vector van het licht in een willekeurige polarisatietoestand altijd kan worden ontbonden. Zo wordt circulaire polarisatie bij voorbeeld gekarakteriseerd door een zuiver imaginaire verhouding van E_x en E_y; een zuiver reële verhouding daarentegen geeft vlakke polarisatie aan onder een hoek arctg (E_y/E_x) met de x-as.

Laat nu E_x en E_y de complexe amplitudes zijn vóór een cyclus, E_x' en E_y' die na een cyclus. Conditie (c) vereist dan:

$$E_y'/E_x' = E_y/E_x. \qquad (16)$$

De geldigheid van conditie (c) is onafhankelijk van de condities (a) en (b) omdat een fase- of een amplitudefactor, die E_x' en E_y' gemeenschappelijk hebben, de geldigheid van (16) niet beïnvloedt. Daarom kunnen, als de spiegels de enige anisotrope elementen in de interferometer zijn, de consequenties van (16) met behulp van de lege interferometer worden bestudeerd. Er zal dan een lineaire relatie bestaan tussen E' en E zodanig dat

$$E_x' - E_x = a_{11}E_x + a_{12}E_y$$

en

$$E_y' - E_y = a_{21}E_x + a_{22}E_y, \qquad (17)$$

waarbij de matrixelementen a_{ij} complexe grootheden zijn. De verhoudingen E_y/E_x die zowel aan (16) als aan (17) voldoen worden gegeven door

$$E_y/E_x = \frac{a_{22}-a_{11}}{2a_{21}} \pm \left[\left(\frac{a_{22}-a_{11}}{2a_{21}}\right)^2 + \frac{a_{21}}{a_{12}}\right]^{\frac{1}{2}} \qquad (18)$$

en definiëren twee, in het algemeen elliptische, polarisatietoestanden, de „polarisatie-eigentoestanden" van de interferometer [6] [5]. Zij worden bepaald door de optische eigenschappen van de spiegels en hangen af van de frequentie; in de beschouwde nieuwe gebiedjes kan deze afhankelijkheid verwaarloosd worden. De spiegels zijn isotroop indien $a_{22} - a_{11} = a_{12} = a_{21} = 0$. In dit geval is E_y/E_x in (18) onbepaald, hetgeen correspondeert met een eerdere opmerking, nl. dat er dan geen beperkende voorwaarde wordt opgelegd aan de polarisatietoestand. In de praktijk zijn $a_{22} - a_{11}$, a_{12} en a_{21} grootheden van de orde van slechts 10^{-4}. Het zijn echter de *verhoudingen* van deze kleine grootheden die, in overeenstemming met (18), de eigentoestanden bepalen.

Opnieuw aannemend dat in een stationair oscillerende laser de rol van het medium alleen bestaat uit het leveren van voldoende versterking om de verliezen goed te maken, kan worden geconcludeerd dat een laser in één of in beide, door de lege interferometer bepaalde, polarisatie-eigentoestanden zou moeten oscilleren. De frequenties die met deze toestanden corresponderen, kunnen als gevolg van de fase-anisotropie van de spiegels verschillend zijn. Omdat van de polarisatie afhankelijke fasesprongen bij de spiegels van de orde van de coëfficiënten a_{ij} zijn (d.w.z. 10^{-4}) en het ordegetal ongeveer 2×10^5 is, verwacht men een relatieve frequentieverschuiving van de orde van $10^{-9}/2\pi$, d.w.z. ongeveer 50 kHz. Het met de twee polarisatietoestanden corresponderende verlies in een lege trilholte kan ook verschillend zijn ten gevolge van absorptie-anisotropie van de spiegels. Gezien het feit dat het frequentieverschil van 50 kHz wel kleiner is dan de inhomogene lijnbreedte γ_{ba}, zodat de twee eigentoestanden verzadigingseffecten van een gemeenschappelijk „burned hole" ondergaan, verwacht men dat alleen die polarisatie-eigentoestand zal verschijnen welke behoort bij het kleinste verlies in de lege trilholte.

Experimentele resultaten laten echter zien dat een laser niet noodzakelijkerwijze in een polarisatie-eigentoestand van de lege interferometer oscilleert en dat het medium niet de tot nu toe veronderstelde neutrale rol speelt. Dit zal later worden besproken. Eerst zal de Poincaré-representatie voor eigentoestanden worden ingevoerd.

De bol van Poincaré

In 1892 introduceerde Poincaré een zeer nuttige methode om polarisatietoestanden aanschouwelijk te maken. Hierbij wordt een toestand gepresenteerd door een punt op een bol, gedefinieerd door een lengte 2ψ en een breedte 2χ (zie *fig. 1*). Het verband tussen een po-

[5] H. de Lang, Proefschrift, Utrecht 1966 (ook gepubliceerd als Philips Res. Repts. Suppl. 1967, no. 8).
[6] H. de Lang, Philips Res. Repts. 19, 429, 1964.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

larisatietoestand en een punt $(2\psi, 2\chi)$ op de bol is dat ψ correspondeert met de hoek tussen de lange as van de polarisatie-ellips en de x-as van de laser, terwijl tg χ de assenverhouding van de ellips is $(-1 \leqslant \text{tg}\ \chi \leqslant 1;\ -\pi/4 \leqslant \chi \leqslant \pi/4)$. In het bijzonder representeren de twee polen circulaire polarisatie met tegengestelde omloopzin; daarentegen is de equator de meetkundige plaats van de lineaire polarisaties. De polarisatie-eigentoestanden die we voorheen besproken hebben, kunnen daarom voorgesteld worden door twee punten ergens op de bol. Als de spiegels alleen fase-anisotropie of alleen absorptie-anisotropie hebben, liggen deze punten diametraal tegenover elkaar.

Met behulp van de voorheen geïntroduceerde complexe amplitudes E_x en E_y laten zich de volgende uitdrukkingen gemakkelijk verifiëren:

$$\begin{aligned} \text{tg}\ 2\psi &= (E_x{}^*E_y + E_y{}^*E_x)/(E_x{}^*E_x - E_y{}^*E_y) \\ \sin 2\chi &= -\,\text{i}(E_x{}^*E_y - E_y{}^*E_x)/(E_x{}^*E_x + E_y{}^*E_y); \end{aligned} \qquad (19)$$

hierin geeft een sterretje de toegevoegd complexe grootheid aan. Merk op dat deze uitdrukkingen alleen functies van (E_y/E_x) of $(E_y/E_x)^*$ zijn.

De bol van Poincaré is ook geschikt voor het aanschouwelijk maken van de evolutie van een polarisatietoestand in de tijd [5]. Laten we, om dit toe te lichten, eens nagaan wat er gebeurt als, op een of ander begintijdstip, de golf in de laser een frequentie ν, een intensiteit I en een zodanige polarisatietoestand heeft dat aan de herhalingscondities van blz. 348 niet exact wordt voldaan. Het spreekt vanzelf dat, weer aangenomen dat alleen de spiegels de polarisatietoestand beïnvloeden, de verandering van E_y/E_x per cyclustijd $E_y'/E_x' - E_y/E_x$ is, en uit vergelijking (17) volgt. De afgeleide naar de tijd van de polarisatietoestand op een gegeven tijdstip zal dan alleen afhangen van de polarisatietoestand op dat tijdstip.

Dit maakt het mogelijk om op de bol van Poincaré banen te construeren die de evolutie beschrijven. Zo'n baan ziet men in *fig. 2*. In het algemeen ligt het beginpunt in de polarisatie-eigentoestand van de lege tril-

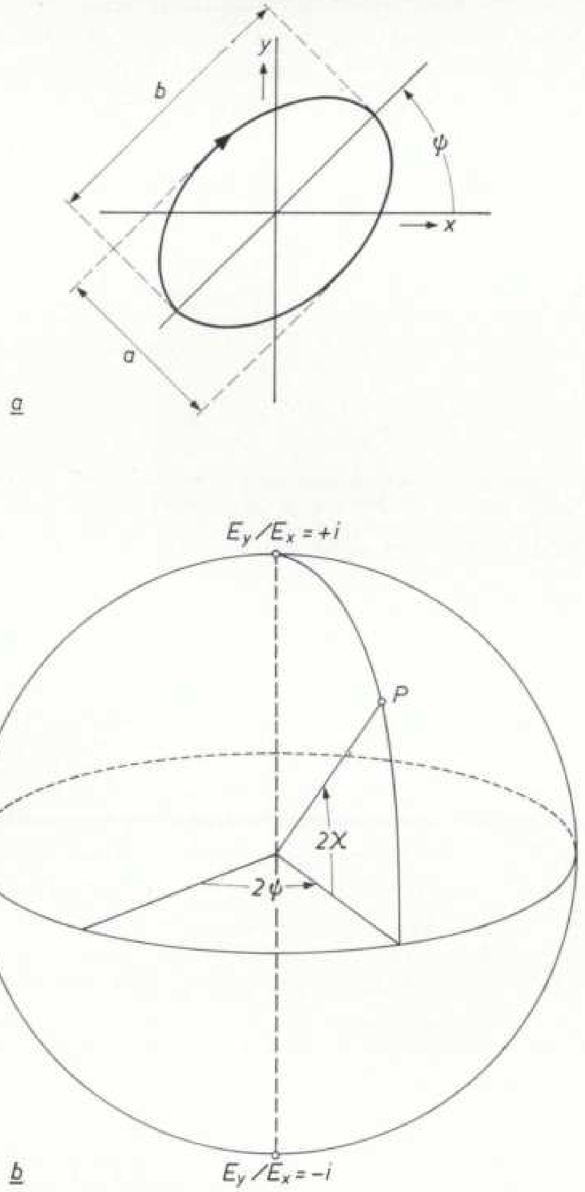


Fig. 1. Een polarisatie-ellips (*a*) en zijn representatie P op de bol van Poincaré (*b*); de lengtecoördinaat op de bol is gelijk aan twee maal het azimut ψ van de lange as en de breedtegraad is 2χ, waarbij tg $\chi = a/b$ $(-1 < \text{tg}\ \chi < +1)$. In het bijzonder is de „equator" $(2\chi = 0)$ de meetkundige plaats van de lineaire polarisatietoestand, terwijl de „polen" $(2\chi = \pm\,\pi/2)$ circulaire polarisatietoestanden met tegengestelde omloopzin representeren.

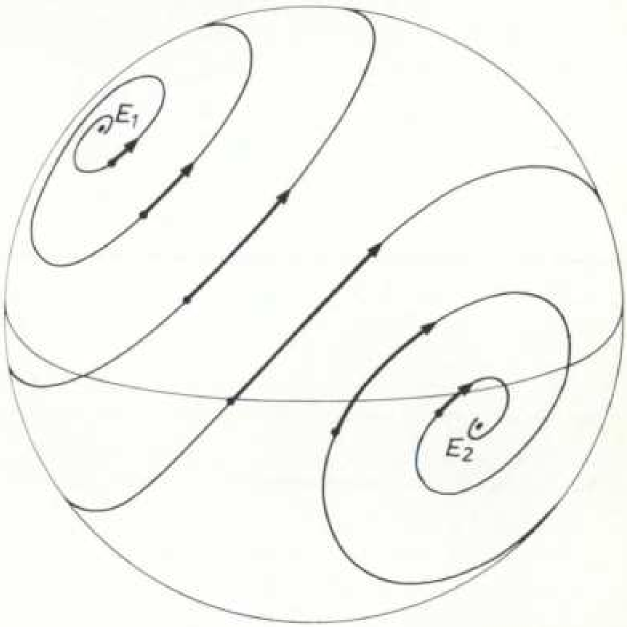


Fig. 2. Een baan op de bol van Poincaré behorende bij een polarisatie van een trillingswijze onder invloed van een willekeurige interferometer-anisotropie. De baan loopt spiraalsgewijze van de eigentoestand E_1 met veel absorptie naar de eigentoestand E_2 met weinig absorptie. De omloopzin van de spiraal wordt bepaald door het teken van het frequentieverschil $\nu_2 - \nu_1$ van de eigentoestanden. In dit voorbeeld is $\nu_2 - \nu_1 > 0$.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

holte met het grootste verlies, en het eindpunt bij de toestand met het kleinste verlies, hetgeen in overeenstemming is met de eerder vermelde opvatting dat deze laatste in een stationaire laser aangetroffen zal worden. De lengte van de pijlen in het stromingsveld in fig. 2 geeft de grootte van de afgeleide naar de tijd aan. Deze is in het onstabiele vertrekpunt en in de stabiele eindpositie gelijk aan nul. In het geval van een laser met alleen absorptie-anisotropie worden de banen halve cirkels die de (diametrale) eigentoestanden met elkaar verbinden; voor zuivere fase-anisotropie zijn het gesloten cirkels die loodrecht staan op de as die de eigentoestanden verbindt. Zo'n geval zullen we in het volgende hoofdstuk tegenkomen (fig. 8).

De geldigheid van deze denkbeelden kan op doorslaggevende wijze getoetst worden aan een experiment [3] [7] [5] waarin een spiegel wordt gebruikt die een deel van één circulaire component van het invallende licht terugkaatst als circulair gepolariseerd licht met tegengestelde draaizin. Van een gewone laserspiegel kan men zo'n bijzondere eigenschap niet verwachten, maar de genoemde reflectie-anisotropie kan wèl verwezenlijkt worden. Daartoe maakt men gebruik van het licht dat een gewone, vrijwel isotrope interferometerspiegel M_1 doorlaat; een deel daarvan wordt, door middel van een uitwendige spiegel M_3, in de interferometer teruggekaatst (*fig. 3*). Tussen M_1 en M_3 bevinden zich in de opstelling nog een polarisator en een $\lambda/4$-plaat.

Voor deze kunstmatig anisotrope spiegel geldt $a_{12} = = ia_{11} = -ia_{22} = a_{21}$, zodat volgens (18) de twee polarisatie-eigentoestanden ontaarden in één enkele, gelegen op de pool waarvoor $E_y/E_x = i$ is. Het daarmee corresponderende stromingspatroon is geschetst in *fig. 4*. Men kan aantonen dat één der banen een meridiaan is die de equator snijdt in een punt met lengte-

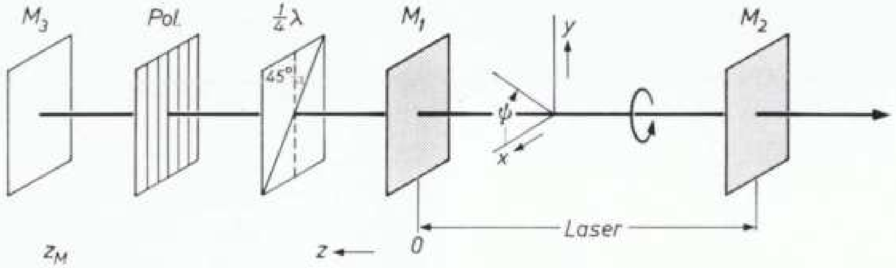


Fig. 3. Een Fabry-Perot-laser met een extern anisotroop reflectiesysteem waardoor de interferometer een ontaarde anisotropie met een circulaire eigentoestand heeft; de polarisatie van de trillingswijze is ten gevolge van verzadigingsanisotropie van het medium vrijwel lineair; het polarisatie-azimut ψ is een maat voor de positie z_M van de externe spiegel M_3 volgens de relatie $\psi = 2\pi z_M/\lambda$ + constante.

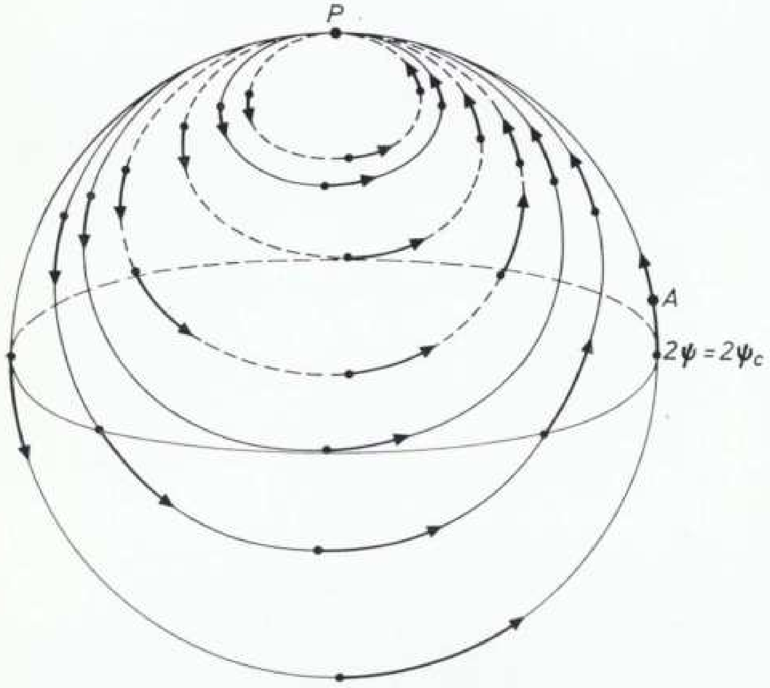


Fig. 4. Stromingspatroon ten gevolge van interferometer-anisotropie in de opstelling van fig. 3. Alle stroomlijnen zijn cirkels die door de circulaire eigentoestand gaan en daar raken aan de meridiaan met azimut $2\psi_c$.

coördinaat $2\ \psi_c = 4\pi z_M/\lambda$ + constante. waarbij z_M de positie van spiegel M_3 is, Deze positie bepaalt de argumenten van de matrixelementen a_{ij}.

Het belangrijke is nu dat experimenteel blijkt dat een laser die uitgevoerd is met dit reflectiesysteem, in een vrijwel lineair gepolariseerde toestand oscilleert, en wel met een azimut ψ dat van z_M afhangt volgens $\psi = 2\pi z_M/\lambda$ + constante. Dit experimentele feit is vanuit verscheidene gezichtspunten interessant. Ten eerste is dit een voorbeeld van een opstelling waarin het polarisatievlak van een lichtbron van grote intensiteit periodiek met een afstand (z_M) varieert. Zo'n opstelling is uit technisch oogpunt van belang omdat het de mogelijkheid opent een apparaat te construeren dat verplaatsingen elektronisch telt en meet in eenheden, gelijk aan de halve golflengte, of zelfs nog kleinere. Het

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

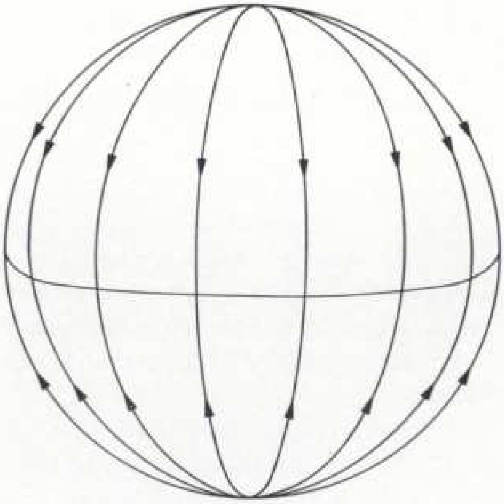


Fig. 5. Patroon van meridionale naar de equator gerichte stroomlijnen ten gevolge van absorptieve verzadigingsanisotropie in een He-Ne-laser, werkend bij een golflengte van 1,152 μm.

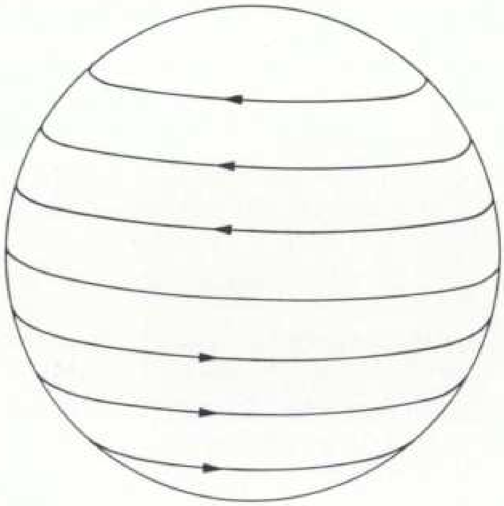


Fig. 6. Patroon van evenwijdig met de equator lopende stroomlijnen ten gevolge van dispersieve verzadigingsanisotropie in een zelfde laser als van fig. 5. De richting van de stroomlijnen wordt bepaald door het teken van ($\nu_{\text{trillingswijze}} - \nu_{\text{lijn}}$) en is in dit voorbeeld positief.

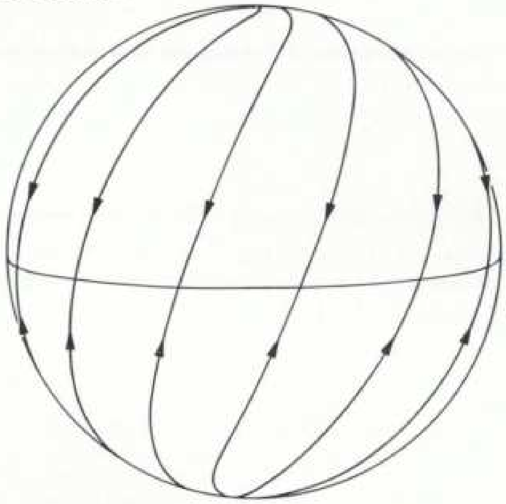


Fig. 7. Stromingspatroon dat ontstaat door combinatie van de effecten van het absorptieve en het dispersieve deel van de verzadigingsanisotropie (superpositie van de stromingspatronen van de figuren 5 en 6).

resultaat is ook om een andere reden verrassend: zoals men in fig. 4 kan zien, is er in de buurt van de equator geen enkel punt waar de afgeleide van de polarisatietoestand naar de tijd nul is. Dat de laserstraling desondanks toch volhardt in een vrijwel lineair gepolariseerde toestand, wettigt de conclusie dat niet alleen de interferometer maar ook het gasvormige medium een rol speelt in de bepaling van de polarisatietoestand.

Nog duidelijker blijkt dit uit de volgende waarnemingen. De polarisatietoestand van de laser is enigszins elliptisch, anders gezegd, hij wordt gerepresenteerd door een punt dat een beetje van de equator afwijkt, en wel in de richting van de verwachte circulair gepolariseerde eigentoestand. De afwijking neemt af met toenemende overmaatfractie $F-1$ boven de drempelwaarde. Dit alles wijst op een effect dat veroorzaakt wordt door de niet-lineaire eigenschappen van het medium. Het lijkt wel alsof er een sterke extra „kracht" aanwezig is die de polarisatietoestand naar de equator drijft, en dat deze kracht groter wordt als de invloed van niet-lineaire of verzadigingseffecten toeneemt. De kracht zou evenredig moeten zijn met de lichtintensiteit in het medium en zou op de bol van Poincaré voorgesteld worden door pijlen die in de polen en op de equator nul zijn en van de polen naar de equator wijzen (*fig. 5*). De laatste twee uitspraken volgen uit symmetrie-overwegingen: het medium zou een voorkeur mogen hebben voor lineaire polarisatie maar het zou noch tussen verschillende waarden van ψ noch tussen de twee mogelijke rotatierichtingen onderscheid mogen maken. Het is, binnen het raam van deze hypothese, mogelijk om in te zien dat de laser een stationaire polarisatietoestand A (zie fig. 4) zal vinden die gelegen is op de meridiaanbaan daar waar de pijlen afkomstig van de anisotropie van de spiegel en van de verzadigingsanisotropie van het medium, elkaar net compenseren. Het is gemakkelijk in te zien dat er een tweede, diametraal tegenover A gelegen stationaire toestand bestaat die, in tegenstelling met A, niet stabiel is. Verder is nu ook duidelijk waarom het azimut ψ op de beschreven wijze van z_M afhangt.

Een nauwkeuriger analyse, die in het volgende hoofdstuk wordt gegeven, leert dat deze interpretatie wel juist maar niet volledig is. We zullen zien dat verzadigingsanisotropie ook aanleiding geeft tot drijvende krachten die een component evenwijdig met de equator hebben; zie het stromingspatroon van *fig. 6*. Zij worden veroorzaakt door het dispersieve deel van de niet-lineaire effecten in het medium. De superpositie van de stromingspatronen uit de figuren 5 en 6 leidt tot het netto, van verzadigingseffecten afkomstige, stromingspatroon dat in *fig. 7* getekend is. In het boven bespro-

[7] H. de Lang, G. Bouwhuis en E. T. Ferguson, Physics Letters **19**, 482, 1965.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

ken experiment veroorzaakt de superpositie een verplaatsing van het stationaire punt A.

Het is interessant om op te merken dat, nog voor ontwikkeling van de quantummechanische theorie van niet-lineaire polarisatie-effecten beschikbaar was, het netto door verzadiging geïnduceerde stromingspatroon volgens fig. 7 gepostuleerd werd om de eerder in dit artikel vermelde polarisatiesprong te kunnen verklaren. In het volgende hoofdstuk zullen we de polarisatiesprong opnieuw bespreken.

Theoretische behandeling

In een meer kwantitatieve beschrijving van polarisatie-effecten wordt het elektromagnetische veld in de laser voorgesteld door een staande golf sin kz, die een hoekfrequentie ω en een langzaam in de tijd variërende amplitude en fase heeft [8] [9] [10] [11]. In het bijzonder wordt het elektrische veld beschreven door:

$$E(z,t) = \tfrac{1}{2}\,[E(t)e^{-i\omega t} + E^*(t)e^{-i\omega t}]\sin kz \qquad (k = p\pi/L). \qquad (20)$$

Hierin is $E(t)$ een langzaam veranderende complexe grootheid. Het doel is nu de bewegingsvergelijkingen voor de tweedimensionele vector $E_x(t)$, $E_y(t)$ te vinden. Omdat de spiegels in een laser discrete elementen zijn, is een dergelijke beschrijving niet zonder meer van toepassing. Zolang echter het effect per cyclus van de spiegels op de golf in de interferometer klein is kan het uitgesmeerd worden over de lengte van de interferometer, zodat de zuiver sinusvormige z-afhankelijkheid in (20) gewaarborgd blijft. De uitsmeringsprocedure bestaat formeel uit het toekennen van anisotrope diëlektrische en absorptieve eigenschappen aan het vacuüm, op een dusdanige wijze dat de van de polarisatie afhankelijke optische weglengte en de door de spiegels veroorzaakte verliezen gemiddeld in rekening gebracht worden. In elk punt wordt dan de totale polarisatie $P(z,t)$ gegeven door de som van een fictieve polarisatie van het vacuüm en de werkelijke diëlektrische polarisatie van het medium.

Voor vrijwel vlakke golven geldt volgens Maxwell:

$$k^2E(z,t) + \frac{1}{c^2}\frac{\partial^2 E(z,t)}{\partial t^2} + \frac{1}{\varepsilon_0 c^2}\frac{\partial^2 P(z,t)}{\partial t^2} = 0, \qquad (21)$$

waarin ε_0 de diëlektrische constante van het vacuüm is. Gebruikmakende van vergelijking (20) krijgen we te maken met de tweede afgeleide van $E(t)\exp i\omega t$, die uit drie termen bestaat. Als we $\omega = ck$ kiezen, valt $k^2E(t)\exp i\omega t$ weg uit (21). De gemengde afgeleide wordt behouden en de term met $\partial^2 E/\partial t^2$ op grond van de veronderstelde langzame verandering van $E(t)$ verwaarloosd. Als voor $P(z,t)$ een met (20) analoge uitdrukking wordt gebruikt, verkrijgt men de vectoriële bewegingsvergelijking

$$\frac{2i\omega}{c^2}\frac{\partial E(t)}{\partial t} = \frac{\omega^2 P(t)}{\varepsilon_0 c^2}, \qquad (22)$$

alsmede de toegevoegd complexe vergelijking. Dit procédé wordt gewoonlijk aangeduid als „adiabatische benadering". In de afleiding van (22) is alleen de grootste term in de tweede afgeleide van $P(z,t)$ behouden. Om de vergelijking te kunnen oplossen moet bekend zijn hoe $P(t)$ van $E(t)$ afhangt. $P(t)$ bestaat uit de som van deze termen.

De eerste daarvan is de fictieve polarisatie van het vacuüm, P_v. Deze hangt lineair samen met E door middel van een tweedimensionele tensor $\tilde{A}$:

$$P_v = \varepsilon_0 \tilde{A} E. \qquad (23)$$

De diëlektrische constante ε_0 van het vacuüm is toegevoegd om $\tilde{A}$ dimensieloos te maken. Het reële deel van $\tilde{A}$ moet zó worden gekozen dat het de fase-anisotropie van de spiegels in rekening brengt en het imaginaire deel de anisotropie [12]. Aangezien vergelijking (17) de verandering van E per cyclustijd $2L/c$ gaf, vinden we, gebruik makend van (22), de volgende relatie tussen A_{ij} en a_{ij}:

$$a_{ij} = -i\omega L A_{ij}/c. \qquad (24)$$

Omdat a_{ij} van de orde 10^{-4} was, is A_{ij} van de orde 10^{-10}. De tweede term is de diëlektrische polarisatie P_1 van het medium in zover deze lineair van E afhangt, d.w.z.: karakteristiek is voor het onverzadigde medium. Gezien de isotropie van het onverzadigde medium geldt de vergelijking

$$P_1 = S_1 \varepsilon_0 E, \qquad (25)$$

waarin S_1 een complexe, sterk van de frequentie afhankelijke scalaire grootheid is. Het (positieve) imaginaire deel van S_1 is evenredig met de (negatieve) absorptiecoëfficiënt α in de omgeving van het centrum ω_{ba} van het Doppler-profiel:

$$\alpha = -(\omega/c)\,\mathrm{Im}\,S_1. \qquad (26)$$

De derde term is de niet-lineaire diëlektrische polarisatie P_3 van het medium, welke voorkomt in de eerste term van de reeksontwikkeling van $P(z,t)$ in oneven machten van $E(z,t)$. Deze term beschrijft niet-lineaire

---

[8] D. Polder en W. van Haeringen, Physics Letters **19**, 380, 1965.
[9] W. van Haeringen, Physics Letters **24A**, 65, 1967.
[10] W. van Haeringen, Phys. Rev. **158**, 256, 1967.
[11] W. van Haeringen en H. de Lang, Phys. Rev. **180**, 624, 1969.
[12] Strikt genomen is dit alleen waar voor reflectiesystemen waarbij geen magnetische materialen of magnetische effecten zijn betrokken. In dat geval is de matrix $\tilde{A}$ symmetrisch. In het algemeen beschrijft het hermitische deel van $\tilde{A}$ de fase-effecten en het antihermitische deel de absorptie-effecten.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

effecten. Als we bedenken dat $\boldsymbol{P}(t)$ dat deel van $\boldsymbol{P}(z,t)$ is dat bij benadering evenredig is met $\exp \mathrm{i}\omega t$ en dat een uitdrukking waarin de derde macht van $\boldsymbol{E}(z,t)$ staat dan alleen evenredig met $\exp \mathrm{i}\omega t$ is als er twee factoren $\boldsymbol{E}(t)$ en één factor $\boldsymbol{E}^*(t)$ in voorkomen, dan volgt:

$$\boldsymbol{P}_3 = \varepsilon_0{}^2 S_2(\boldsymbol{E}\cdot\boldsymbol{E}^*)\boldsymbol{E} + \varepsilon_0{}^2 S_3(\boldsymbol{E}\cdot\boldsymbol{E})\boldsymbol{E}^*. \qquad (27)$$

De bijzondere vorm van (27) met de twee complexe scalaire grootheden S_2 en S_3 vloeit voort uit de eis dat de uitdrukking invariant moet zijn ten aanzien van de keuze van de coördinaatassen x en y, met andere woorden: vloeit voort uit de rotatiesymmetrie van het lasermedium. Merk verder op dat er geen kwadratische termen in $\boldsymbol{E}$ en $\boldsymbol{E}^*$ optreden. Dat is een gevolg van de inversiesymmetrie van het medium.

De bewegingsvergelijking luidt nu als volgt:

$$\frac{2\mathrm{i}}{\omega}\frac{\partial \boldsymbol{E}}{\partial t} = \tilde{A}\boldsymbol{E} + S_1\boldsymbol{E} + S_2\varepsilon_0(\boldsymbol{E}\cdot\boldsymbol{E}^*)\boldsymbol{E} + S_3\varepsilon_0(\boldsymbol{E}\cdot\boldsymbol{E})\boldsymbol{E}^*. \qquad (28)$$

Hier is het nuttig op te merken dat op elk tijdstip de vector in het rechterlid kan worden ontbonden in een vector $\boldsymbol{P}_{\parallel}/\varepsilon_0$ evenwijdig met $\boldsymbol{E}$ en een vector $\boldsymbol{P}_{\perp}/\varepsilon_0$ loodrecht op $\boldsymbol{E}$. Alleen de laatste kan een verandering in de polarisatietoestand bewerkstelligen. Dit houdt in dat de termen met S_1 en S_2 niet bijdragen tot een dergelijke verandering. Van de component $\boldsymbol{P}_{\parallel}$, waaraan alle vier termen in (28) bijdragen, veroorzaakt het uitfase-deel $\boldsymbol{P}_{\parallel,\mathrm{out}}$ een verandering in de intensiteit en het in-fase-deel $\boldsymbol{P}_{\parallel,\mathrm{in}}$ een correctie op de frequentie.

De afgeleiden naar de tijd van de polarisatieparameters ψ en χ kunnen met de vergelijkingen (19) en (28) worden berekend. Men vindt dan:

$$\frac{1}{\omega}\frac{\partial\chi}{\partial t} = \Gamma_1(\psi,\chi) - (\mathrm{Im}\, S_3)\,(I/4)\sin 4\chi, \qquad (29a)$$

$$(\cos 2\chi)\,\frac{1}{\omega}\frac{\partial\psi}{\partial t} = \Gamma_2(\psi,\chi) - (\mathrm{Re}\, S_3)(I/4)\sin 4\chi. \qquad (29b)$$

Γ_1 en Γ_2 hangen alleen van de matrixelementen van $\tilde{A}$ af en beschrijven interferometer-anisotropieën. Voor isotrope interferometers geldt: $\Gamma_1 = \Gamma_2 = 0$. De grootheid I is evenredig met de intensiteit en is gedefinieerd door:

$$I = \varepsilon_0(\boldsymbol{E}\cdot\boldsymbol{E}^*). \qquad (30)$$

De afgeleide naar de tijd van I zelf volgt ook uit (28):

$$\frac{1}{\omega I}\frac{\partial I}{\partial t} = \Gamma_3(\psi,\chi) + \mathrm{Im}\, S_1 + I[\mathrm{Im}\,(S_2 + S_3) - (\mathrm{Im}\, S_3)\sin^2 2\chi]. \qquad (31)$$

De grootheid Γ_3 hierin hangt af van dat deel van de matrix $\tilde{A}$ dat de verliezen beschrijft. Voor isotrope interferometers is het een negatieve scalar ten bedrage van $c/2\omega L$ maal het trilholteverlies $\mathcal{L}$ per cyclus, d.w.z. het is van de orde van grootte van 10^{-8}. In een lege anisotrope interferometer is de gemiddelde verzwakking $\Gamma_{\mathrm{av}} = \mathrm{Im}\,(A_{11} + A_{22})/2 < 0$. Dit is in absolute waarde altijd groter dan het van ψ en χ afhangende deel van Γ_3. In een laser wordt Γ_{av} overgecompenseerd door de positieve term $\mathrm{Im}\, S_1$, die de onverzadigde versterking van het medium representeert. De overmaatfractie $F - 1$, die in een representatief geval ongeveer 10% bedraagt, wordt gegeven door

$$F - 1 = \frac{\mathrm{Im}\, S_1 - |\Gamma_{\mathrm{av}}|}{|\Gamma_{\mathrm{av}}|}, \qquad (32)$$

waarbij voor $\mathrm{Im}\, S_1$ de bij de laserfrequentie ω_l behorende waarde moet worden genomen. De term in (31), ten slotte, die evenredig is met I, is altijd negatief en stelt de door verzadiging veroorzaakte vermindering in versterking voor, welk effect, zoals in een vorig hoofdstuk werd besproken, stationair oscilleren mogelijk maakt. Zijn orde van grootte is $(F-1)|\Gamma_{\mathrm{av}}|$, d.w.z. 10^{-9}. Daaruit volgt dat in de in dit artikel besproken experimenten de verzadigingseffecten een orde van grootte groter zijn dan interferometer-anisotropieën.

De termen in (29) die evenredig zijn met I, representeren de ten gevolge van verzadigingsanisotropie op de polarisatietoestand werkende „drijvende krachten". In het bijzonder levert de term die $\mathrm{Im}\, S_3$ bevat het stromingspatroon van fig. 5 en de term die $\mathrm{Re}\, S_3$ bevat het patroon van fig. 6: hun superpositie is in fig. 7 te zien.

De stationaire toestand waarin een laser uiteindelijk zal oscilleren, wordt bepaald door die waarden van I, ψ en χ waarvoor alle afgeleiden naar de tijd in (29) en (31) nul zijn. Door hun relaties zijn deze vergelijkingen zeer gecompliceerd. In veel praktische gevallen, d.w.z. voor niet al te kleine waarden van $F - 1$, zijn de verzadigingseffecten dominant. Voor de helium-neonlaser bij $\lambda = 1{,}152$ µm betekent dit dat χ klein is, zodat I wel constant kan worden genomen. De waarde van I vinden we uit (31) met $\partial I/\partial t = 0$ en $\chi = 0$:

$$I \approx \frac{\mathrm{Im}\, S_1 - |\Gamma_{\mathrm{av}}|}{|\mathrm{Im}\,(S_2 + S_3)|} \approx \frac{F-1}{F}\,\frac{\mathrm{Im}\, S_1}{|\mathrm{Im}\,(S_2 + S_3)|}. \qquad (33)$$

De bespreking van (29) wordt, als I constant is, aanmerkelijk vereenvoudigd.

Alle constanten in (28) kunnen in principe worden gemeten. De spiegelanisotropieën kunnen worden gecalibreerd en de waarden van S_1, S_2 en S_3 kunnen worden gevonden door een zorgvuldige bestudering van de wijze waarop de stationaire polarisatietoestand van de intensiteit afhangt. De grootheden S_1, S_2 en S_3 kunnen ook berekend worden met behulp van een quantummechanische behandeling van de responsie van het medium op een elektromagnetisch veld[10]. Een dergelijke

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

berekening levert tevens de frequentie-afhankelijkheid; de imaginaire (absorptieve) delen van S_1, S_2 en S_3 zijn even, de reële (dispersieve) delen oneven functies van $(\omega - \omega_{ba})$, waarbij ω_{ba} het centrum van het Doppler-profiel is.

Interpretatie van enkele experimenten

De theorie zal nu gebruikt worden om de al eerder genoemde polarisatiesprong te verklaren, uitgaande van een kleine fase-anisotropie van de spiegels, die bij voorbeeld afkomstig is van dubbele breking ten gevolge van spanning in de diëlektrische lagen op de spiegels en van verzadigingsanisotropie van het medium. Zonder aan algemeenheid in te boeten, kunnen we voor de x-as de anisotropie-as van de trilholte nemen; daardoor wordt de matrix $\bar{A}$ gereduceerd tot een vorm waarin $A_{11} = -A_{22} = a$, $A_{12} = A_{21} = 0$. Men kan dan afleiden dat

$$\left.\begin{array}{l} \Gamma_1 = \tfrac{1}{2}a \sin 2\psi, \\ \Gamma_2 = -\tfrac{1}{2}a \cos 2\psi \sin 2\chi, \\ \Gamma_3 = 0. \end{array}\right\} \qquad (34)$$

Het corresponderende stromingspatroon bestaat uit cirkels, zoals weergegeven wordt in *fig. 8*. Door de term met de positief veronderstelde term Im S_3 in vergelijking (29a) wordt de polarisatietoestand naar de equator gedreven en zal de uiteindelijke stationaire toestand lineair gepolariseerd zijn met $\chi = 0$ en $2\psi = 0$ of π. Voor deze waarden zijn $\partial\psi/\partial t$ en $\partial\chi/\partial t$ nul. Er zijn dus twee oplossingen, waarvan de ene (met $\psi = 0$) gepolariseerd is in de x-richting, de andere (met $\psi = \pi/2$) in de y-richting. De corresponderende

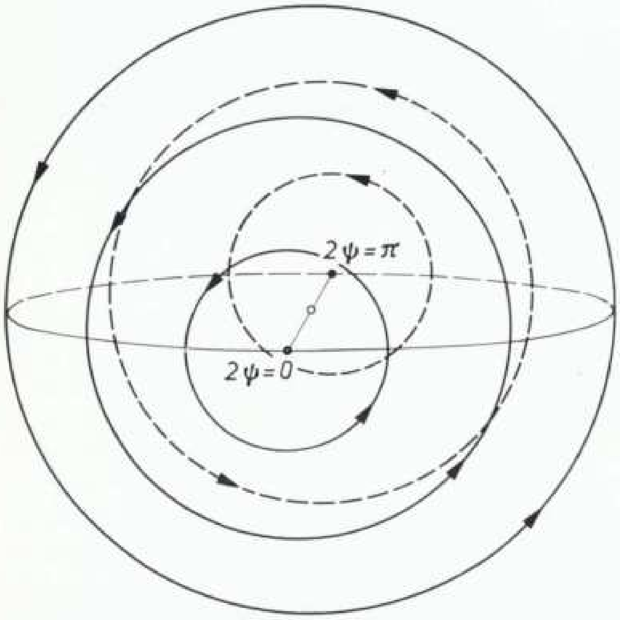


Fig. 8. Stroomlijnen ten gevolge van de fase-anisotropie van de interferometer. De eigentoestanden zijn lineair gepolariseerd, die met de laagste frequentie bij $\psi = 0$, en die met de hoogste frequentie bij $\psi = \pi/2$.

stationaire waarden I_0 van de intensiteit zijn aan elkaar gelijk, hetgeen in overeenstemming is met het afwezig zijn van absorptie-anisotropie. De vraag is welke oplossing stabiel is.

Om dit te onderzoeken beschouwen we de superpositie van het stromingspatroon van de interferometer-anisotropie (fig. 8) en dat van de door verzadiging geïnduceerde anisotropie (fig. 7) bij een waarde van I gegeven door (33). In *fig. 9* is het resulterende stromingspatroon in een klein gebied rond het stationaire punt ($\chi = 0$, $\psi = 0$) getekend. Evenzo zien we in *fig. 10* de omgeving van het andere stationaire punt ($\chi = 0$, $2\psi = \pi$). De figuur met de convergerende stroomlijnen correspondeert met de stabiele toestand, terwijl de divergerende stroomlijnen op instabiliteit wijzen.

Nader onderzoek van de figuren 9 en 10 leert dat de helling van de door verzadiging geïnduceerde stroomlijnen ten opzichte van de meridiaan (zie fig. 7) in de boven gegeven beschouwingen een essentiële rol speelt. De helling hangt af van de term in vergelijking (29b) die Re S_3 bevat. Een verandering in teken van deze term zou een helling in tegengestelde richting en daarmee een verwisseling van stabiele en instabiele punten veroorzaken. Dit is nu precies de polarisatiesprong. Omdat Re S_3 een oneven functie van $(\omega - \omega_{ba})$ is, brengt verstemming van de laser van $\omega > \omega_{ba}$ naar $\omega < \omega_{ba}$ deze tekenverandering met zich mee en doet de laser, met een abrupte verandering in azimut van $\pi/2$, van de ene polarisatietoestand naar de andere „springen".

In de buurt van het sprongpunt — daar is $\omega \approx \omega_{ba}$ en dus Re $S_3 \approx 0$ — treden er interessante neveneffecten op. Een verklaring van deze effecten, bistabiliteit en hysterese, kan ook met behulp van de algemene theorie worden gegeven.

De analyse die tot vergelijking (28) leidt, kan worden uitgebreid tot het geval waarin een constant uitwendig magnetisch veld B wordt aangelegd. Als we ons beperken tot een axiaal veld, d.w.z. een veld evenwijdig met de z-richting, dan vinden we extra-termen. Voor kleine B wordt vooral de relatie tussen $\boldsymbol{P}_1$ en $\boldsymbol{E}_1$ belangrijk beïnvloed, hoewel er daarnaast ook termen in de niet-lineaire responsie van het medium verschijnen die van B afhangen. Het belangrijkste effect is de Faraday-rotatie in het medium: het magnetisch veld veroorzaakt een verschil tussen de voortplantingssnelheden van circulair gepolariseerd licht van tegengestelde rotatierichting. Bovendien kan een verschil in (negatieve) absorptiecoëfficiënt ontstaan, dat verwaarloosd zal worden.

Nu is S_1 geen scalaire grootheid meer en wordt (25):

$$\boldsymbol{P}_1 = \bar{b}\boldsymbol{E}, \qquad (35)$$

waarin $\bar{b}$ een antisymmetrische tensor ($b_{12} = -b_{21}$) is

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

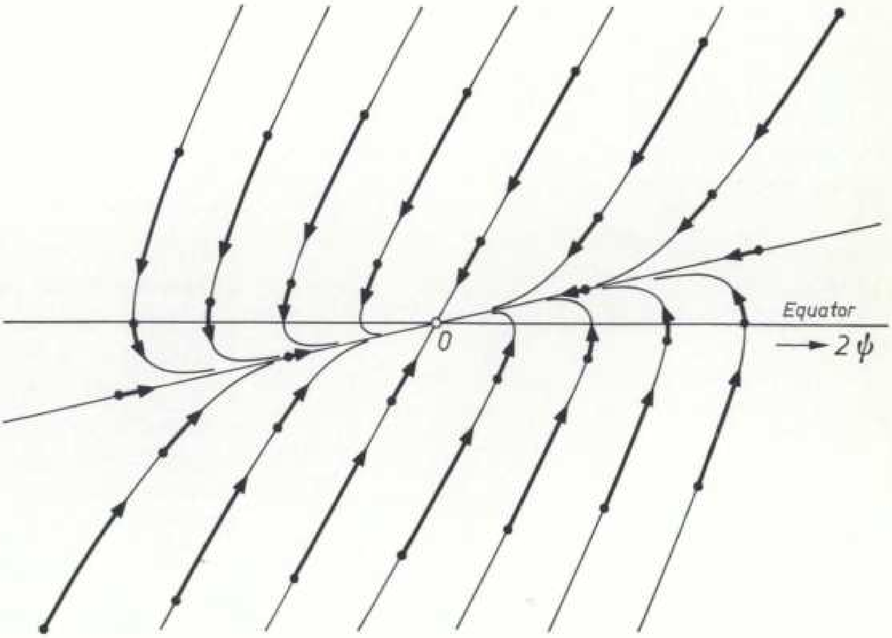


Fig. 9. Superpositie van het stromingspatroon van verzadigingsanisotropie (fig. 7) en dat van interferometer-anisotropie fig. 8), weergegeven in een klein gebied rond het punt $2\psi = 0$, $2\chi = 0$. De convergentie van de stroomlijnen bevestigt de experimenteel waargenomen stabiliteit van dit punt. De lengte van de pijlen correspondeert voor onze experimenten met ca. 1 μs.

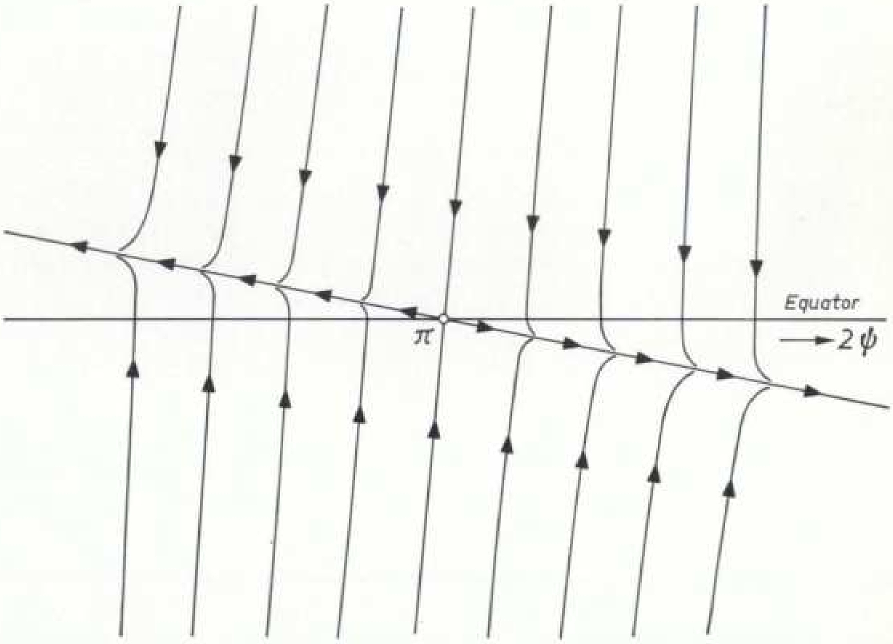


Fig. 10. Superpositie van het stromingspatroon van verzadigingsanisotropie (fig. 7) en dat van interferometer-anisotropie (fig. 8), weergegeven in een klein gebied rond het punt $2\psi = \pi$, $2\chi = 0$. Uit de divergentie van de stroomlijnen volgt dat dit punt niet stabiel is.

met $b_{11} = b_{22}$. In het bijzonder is b_{12} voor kleine waarden van B evenredig met B. Het imaginaire deel β daarvan beschrijft het Faraday-effect. Het stromingspatroon dat correspondeert met dit type anisotropie in een lineair medium, bestaat uit cirkels evenwijdig met de equator, alle met dezelfde omloopzin (*fig. 11*). In (29b) verschijnt een extra-term

$$\frac{1}{\omega}\frac{\partial\psi}{\partial t} = -\frac{\beta}{2}. \qquad (36)$$

Het is duidelijk dat de stationaire toestanden van een isotrope laser in een magnetisch veld worden voorgesteld door de twee, met circulaire polarisatie corresponderende, polen op de bol van Poincaré. De aanwezigheid echter van verzadigingsanisotropie, zoals weergegeven in fig. 7, maakt deze punten instabiel. In feite is er geen stabiele polarisatietoestand. De polarisatietoestand van de laser zal uiteindelijk een eenparige rotatie langs de equator uitvoeren. De toestand zal lineair gepolariseerd zijn met een azimut dat lineair van de tijd afhangt. Achter een in de bundel geplaatste polarisator zal een in de tijd gemoduleerde intensiteit worden waargenomen met een modulatiefrequentie die evenredig is met B; voor $B = 1$ gauss is dat ca. 50 kHz,

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

hetgeen correspondeert met een waarde van 10^{-10} voor β. In deze situatie krijgt de equatoriale stroomlijn in fig. 11 een zeer reële fysische betekenis.

Als er naast het magnetische veld ook een kleine fase-anisotropie in de interferometer aanwezig is (zie fig. 8), treden er interessante effecten op. De vergelijkingen (29) met de extra-term (36), gecombineerd met (34) kunnen bij benadering worden opgelost. Als de verzadigingsanisotropie dominant is, d.w.z. als zowel Im S_3 als Re S_3 veel groter is dan a/I en β/I, vinden we:

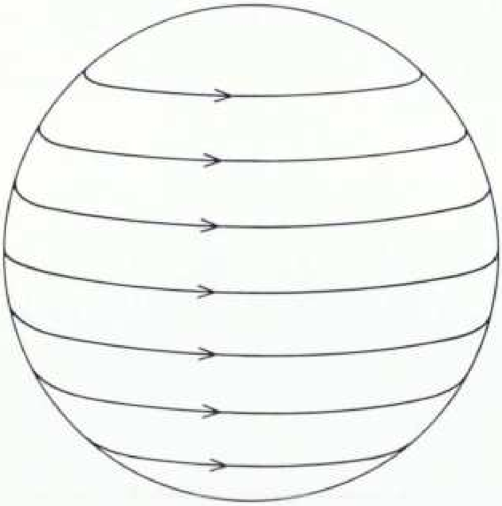


Fig. 11. Stromingspatroon ten gevolge van de door een longitudinaal magnetisch veld veroorzaakte Faraday-rotatie in het lasermedium.

$$\chi \approx (2I \operatorname{Im} S_3)^{-1}\, a \sin 2\psi \ll 1, \tag{37a}$$

$$\frac{1}{\omega}\frac{\partial\psi}{\partial t} \approx -(2 \operatorname{Im} S_3)^{-1}\, a \operatorname{Re} S_3 \sin 2\psi - \tfrac{1}{2}\beta. \tag{37b}$$

Oplossing (37a) leert dat de polarisatietoestand langs een ecliptica loopt die door de eigentoestanden van de interferometer-anisotropie gaat en een kleine helling met de equator maakt (*fig. 12*). Vergelijking (37b) leert dat op deze ecliptica de dispersieve verzadigingsanisotropie Re S_3 zich periodiek laat voelen en de „Faraday"-snelheid $\frac{1}{2}\omega\beta$ afwisselend doet toenemen en afnemen. Dit heeft tot gevolg dat de beweging langs de ecliptica niet meer eenparig is. Terwijl de intensiteit I nagenoeg constant blijft, zal de bundelintensiteit die via een polarisator wordt waargenomen, een niet-sinusvormige functie van de tijd zijn (*fig. 13*).

Voor voldoend grote fase-anisotropie a of voldoend kleine Faraday-constante β zal de continue beweging langs de ecliptica worden afgebroken en wordt een stationaire toestand mogelijk, waarvoor dus $\partial\psi/\partial t = 0$. De laser is dan vrijwel lineair gepolariseerd met een azimut ψ, dat zeer gevoelig is voor veranderingen in de verhouding β/a. Al deze effecten zijn waargenomen en verklaard [3] [4] [5] [9] [10] [11]. In feite waren het deze experimenten met een magnetisch veld die ons er toe brachten het begrip verzadigingsanisotropie in te voeren.

Quantummechanische beschouwingen

De fysica achter de door verzadiging geïnduceerde anisotropie berust op het gedrag van individuele atomen, dat alleen met quantummechanica adequaat beschreven kan worden. Het is echter zeer wel mogelijk in kort bestek duidelijk te maken wat er gebeurt. Beschouw daartoe een atoom waarvan de toestand (b) van hogere energie en de toestand (a) van lagere energie beide de waarde $\frac{1}{2}$ hebben voor het quantumgetal j van het totale impulsmoment. In de helium-neon-laser ligt de situatie anders, maar daar zullen we later op terugkomen. Beide toestanden zijn tweevoudig ontaard en de subniveaus kunnen onderscheiden worden door de projectie van j op de z-as, d.w.z. we hebben toestanden $a_{\frac{1}{2}}$, $a_{-\frac{1}{2}}$ en $b_{\frac{1}{2}}$, $b_{-\frac{1}{2}}$. Uit eenvoudige symmetriebeschouwingen volgt dat door zich in de z-richting voortplantend licht alleen optische overgangen tussen de paren $a_{\frac{1}{2}} \leftrightarrow b_{-\frac{1}{2}}$ en $a_{-\frac{1}{2}} \leftrightarrow b_{\frac{1}{2}}$ geïnduceerd kunnen worden, waarbij het eerste paar alleen in wisselwerking treedt met de, laten we zeggen, rechtsdraaiende circulaire component van het licht, en het laatste paar alleen met de linksdraaiende circulaire component. De (negatieve) absorptiecoëfficiënten van de rechtsdraaiende en de linksdraaiende component worden dientengevolge bepaald door respectievelijk de relatieve overbezetting van $b_{-\frac{1}{2}}$ ten opzichte van $a_{\frac{1}{2}}$ en van $b_{\frac{1}{2}}$ ten opzichte van $a_{-\frac{1}{2}}$.

Stel dat op een begintijdstip een elliptische polarisatietoestand bestaat waarin de rechtsdraaiende circulaire component een grotere amplitude heeft dan de linksdraaiende. De verzadiging veroorzaakt dan een grotere reductie in de overbezetting van het paar $a_{\frac{1}{2}} \leftrightarrow b_{-\frac{1}{2}}$ dan in die van het andere paar. Daardoor zal de linksdraaiende component ten opzichte van de rechtsdraaiende component aangroeien en de elliptische polarisatietoestand meer op een lineaire gaan lijken. Dit is in essentie de verklaring van de langs de equator gerichte, door verzadiging geïnduceerde drijvende kracht op de bol van Poincaré voor een overgang $j = \frac{1}{2} \leftrightarrow j = \frac{1}{2}$. In de laser met $\lambda = 1{,}152$ µm is er sprake van een overgang tussen een drievoudig ontaard niveau met $j = 1$ en een vijfvoudig ontaard niveau met $j = 2$. We hebben daarom een theorie [8] ontwikkeld voor het berekenen van de waarden van S_1, S_2 en S_3 voor mogelijke paren van geheelwaardige of halfwaardige (j, j') die aan de selectieregel $|j' - j| = 0$ of 1 voldoen.

Niet geheel onverwacht blijkt dat de waarden van S_1, S_2 en S_3 evenredig zijn met $N_b - N_a$. Helaas is dit een

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

grootheid die sterk van de experimentele omstandigheden afhangt en die zich niet direct laat meten. De verhoudingen S_2/S_1, S_3/S_1 of (Im S_3)/(Im S_1) zullen deze grootheid echter *niet* bevatten. Verder kan het imaginaire deel van S_1 met behulp van vergelijking (32) worden uitgedrukt in meetbare grootheden. Het is gelijk aan $F|\Gamma_{av}|$. Derhalve kan ook Im S_3 worden uitgedrukt in termen van F en $|\Gamma_{av}|$. Een uitvoerige berekening leert:

$$\mathrm{Im}\, S_3 = \frac{F|\Gamma_{av}|}{8\hbar^2\varepsilon_0}\frac{p_{ab}{}^2}{\gamma_a\gamma_b}\left[\frac{(\omega-\omega_{ba})^2+2\gamma_{ab}{}^2}{(\omega-\omega_{ba})^2+\gamma_{ab}{}^2}\right]g(j), \quad (38)$$

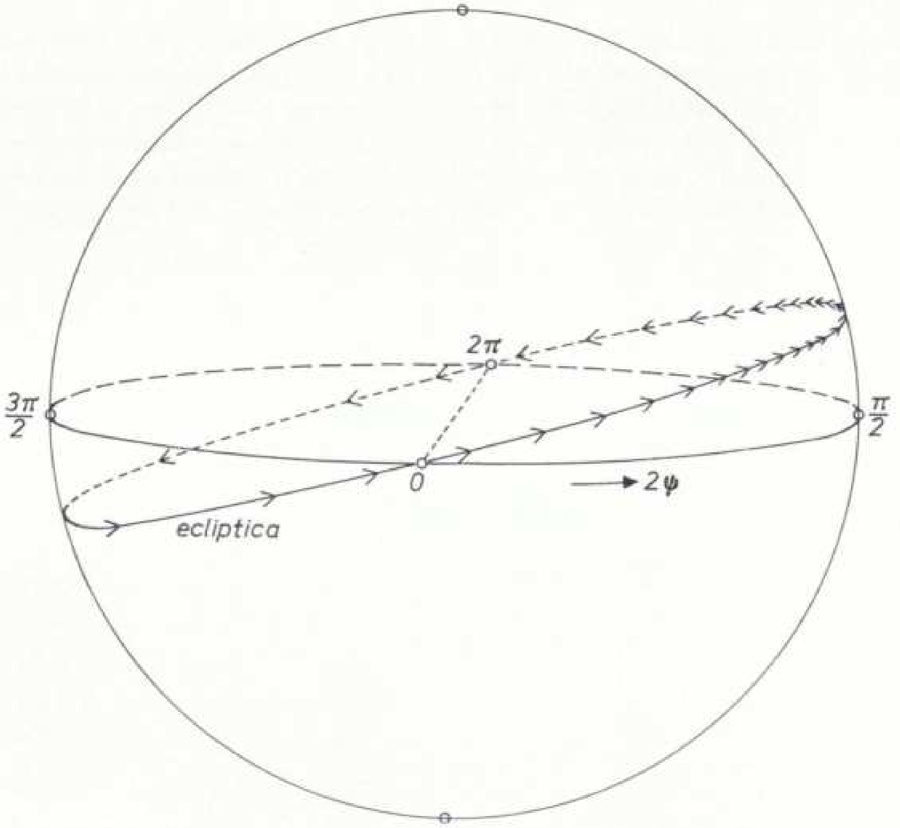


Fig. 12. Periodieke beweging van de polarisatietoestand langs een ecliptica. De samenstellende stromingspatronen zijn die van de figuren 7, 8 en 11.

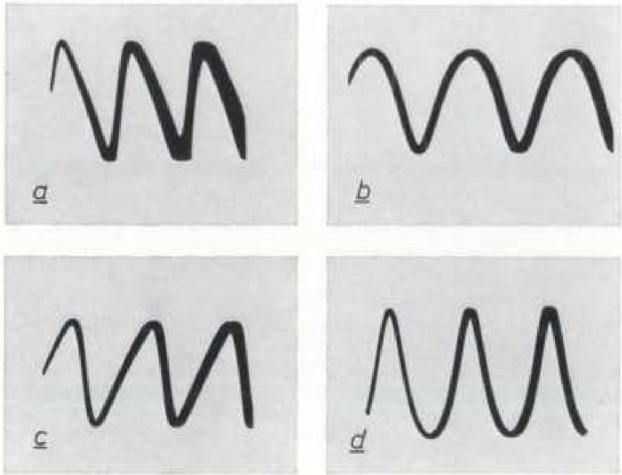


Fig. 13. Oscillogrammen waaruit de niet-eenparige rotatie van het polarisatie-azimut volgens fig. 12 blijkt. De laserbundel werd door een polarisator gestuurd, waarna de doorgelaten intensiteit als functie van de tijd oscillografisch werd opgenomen (de lineaire tijdas was ongeveer 3×10^{-4} s). Voor de oscillogrammen (*a*), (*b*), (*c*) en (*d*) was het azimuth ψ van de door de polarisator doorgelaten trillingsrichting respectievelijk 0, $\pi/4$, $\pi/2$ en $3\pi/4$. De golfvorm voor elk van deze gevallen bevestigt de in fig. 12 geschetste snelheidsmodulatie langs de ecliptica.

waarin p_{ab} het matrixelement van de elektrische dipool tussen de boven- en de benedentoestand is en $g(j)$ gedefinieerd is door

$$g(j) = -\frac{1}{5}(j-1)(j+2) \text{ voor een overgang } j \leftrightarrow j,$$
$$g(j) = +\frac{2}{5}j(j+2) \text{ voor een overgang } j \leftrightarrow j+1. \quad (39)$$

Op gelijksoortige wijze kunnen uitdrukkingen worden verkregen voor S_3 zelf en voor S_2.

De vergelijkingen (38) en (39) zijn om de volgende redenen bijzonder interessant. Tot nu toe is impliciet verondersteld dat Im S_3 een positieve grootheid was die aanleiding gaf tot door verzadigingsanisotropie veroorzaakte preferentie voor lineaire polarisatie. Inderdaad volgt uit (39) zowel voor de neonovergang $\lambda = 1{,}152$ µm, $j = 1 \leftrightarrow j = 2$ als voor de boven beschouwde overgang $j = \frac{1}{2} \leftrightarrow j = \frac{1}{2}$ dat Im S_3 positief is. Maar dat geldt niet voor alle overgangen. Uit (39) volgen de regels [8]:

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

$$\text{Im } S_3 > 0 \text{ voor } \tfrac{1}{2} \leftrightarrow \tfrac{1}{2} \text{ en alle } j \leftrightarrow j+1 \ (j > 0), \quad (40a)$$

$$\text{Im } S_3 < 0 \text{ voor } j \leftrightarrow j \ (j > 1), \quad (40b)$$

$$\text{Im } S_3 = 0 \text{ voor } 0 \leftrightarrow 1 \text{ en } 1 \leftrightarrow 1. \quad (40c)$$

Dit frappante resultaat van de theorie voorspelt drie typen van verzadigingsanisotropie. Die kan leiden òf tot voorkeur voor lineaire polarisatie (geval 40a), òf tot voorkeur voor circulaire polarisatie (geval 40b), d.w.z. gelijke voorkeur voor de polen op de bol van Poincaré. Ten slotte kan verzadigingsanisotropie geheel onwerkzaam zijn (geval 40c), in welk geval het dispersieve deel Re S_3 óók nul is.

Deze theoretische voorspellingen waren voor ons aanleiding een overgang $\lambda = 1{,}207$ µm, $j = 2 \leftrightarrow j = 2$ in de helium-neon-laser te onderzoeken. De experimenten toonden aan dat er in dit geval inderdaad een sterke voorkeur bestaat voor circulaire polarisatie en dat er inderdaad een met de twee polen op de bol van Poincaré corresponderende bistabiliteit optreedt [13] [5].

Experimenten [13] [5] met de neonovergang $\lambda = 1{,}523$ µm, $j = 1 \leftrightarrow j = 0$ bevestigden echter niet het door de theorie voorspelde neutrale gedrag. In plaats daarvan werd een duidelijke voorkeur voor circulaire polarisatie waargenomen, maar wel zwakker dan in de gevallen van het type (40b). Dit resteffect kon worden verklaard [14] [15] met behulp van een tot nu toe niet beschouwd mechanisme, nl. het relaxatieproces tussen de niveaus van een ontaarde ($j = 1$) toestand. We konden aantonen dat de waargenomen voorkeur ontstaat als het werkzame botsingsproces van zodanige aard is dat de impulsmomentrelaxatie van het ontaarde niveau sneller plaatsvindt dan de quadrupoolrelaxatie. Er zij overigens op gewezen dat anderen, weliswaar in een heel ander verband en in een ander atoom, reeds een verschil tussen de twee relaxatiesnelheden gevonden hebben.

In dit artikel hebben wij ernaar gestreefd een naar onze mening coherent beeld van het onderwerp te geven. In de aangehaalde artikelen kan men referenties naar belangrijk werk van andere onderzoekers op dit gebied vinden.

[13] H. de Lang en G. Bouwhuis, Physics Letters **20**, 383, 1966.
[14] D. Polder en W. van Haeringen, Physics Letters **25A**, 337, 1967.
[15] G. Bouwhuis, Physics Letters **27A**, 693, 1968.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

On Electromagnetic Units

by H. B. G. Casimir

(4 III 68)

Once upon a time there was in a far away country a great, great kitchen in which many cooks plied their trade and in which there was a great profusion of pots and pans and kettles and cauldrons and bowls and basins of every size and kind and description. Some of these vessels were empty but others contained eggs or rice or apples or spices and many other delectable things. Now the cooks, if they were not busy broiling and baking and cooking and frying and preparing sundry soups and sauces, amused themselves with philosophical speculation and so it came to happen that the art of tagenometry (from ταγηνον, a frying-pan) was developed to great perfection. Sometimes it was even referred to as panmetry, the art of measuring everything, but the ignorant scullions, misinterpreting the word, promptly also spoke about potmetry, much in the same way in which the transatlantic chefs have supplemented the hamburger with a cheeseburger.

To every vessel tagenometry assigned a volume V. This was measured in cubic inches and determined by measuring dimensions with great precision and by then applying the formulae of solid geometry or in case of irregular shapes by numerical integration on a beanbeaded abacus. But to every vessel there was also assigned an entirely different quantity, the volumetric displacement W. This was measured in gallons and determined by filling the vessel with water, pouring out the water, weighing said water in pounds avoirdupois, correcting for temperature and dividing by 10. The ratio of volumetric displacement and volume was referred to as the volumetric constant, $\varepsilon = W/V$. In the course of time it became clear that this volumetric constant had the same value for every empty vessel; this became known as the volumetric constant of empty space or ε_0. But for other vessels the volumetric constant behaved often in an erratic way. It changed after thermal treatment, or simply with time, it depended on the speed of measurement. Also the dynamic behaviour of moving non-empty pans posed curious problems.

One day a wise man entered the kitchen and after having listened to the worried cooks he said: 'I can solve your problems. There is really only one tagenometric quantity, let us call it the volume and measure it in cubic centimeters. Weighing water will give the same value for an empty vessel if you take the weight in grams. So your volumetric constant of empty space is just unity. But in a non-empty pan part of the volume is occupied by edibles like potatoes or pears or plums; let us call this volume P. Then, with the water-method you determine $V - P$. In many cases P will be proportional to V, that is $P = \varkappa V$. Then the water-weight volume, your volumetric displacement is $W = V - \varkappa V = (1 - \varkappa) V$ and hence $\varepsilon = 1 - \varkappa$. What you really should study is P and its dependance on the constitution and preparation of the victuals. And instead of studying the dynamics of a non-empty pan, you should study the motion of the things it contains.'

Philips PD-ultra-microfilm

Terugvergroting van een micro-opname op PD-film (verkleining 500×) van een bladzijde uit Helvetica Physica Acta. *(In het volgende nummer van Philips technisch Tijdschrift zal een artikel over het PD-proces worden opgenomen, Red.)*

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

 Philips techn. T. 32, 360-363, 1971, no. 10/11/12

De temperatuurverdeling in met gas gevulde gloeilampen

J. Fitzgerald en H. Hörster

Belangrijke fysische eigenschappen van met gas gevulde gloeilampen, zoals levensduur en rendement, worden bepaald door de massa- en de warmtestroming aan het oppervlak van de gloeidraad. Voor het beschrijven van deze processen moet men de temperatuur- en snelheidsverdeling van het edelgas in de ballon nauwkeurig kennen. In principe kan men deze verdelingen berekenen met behulp van de bewegingsvergelijkingen voor vloeistoffen (Navier-Stokes), de energie- en de continuïteitsvergelijking. Door wiskundige moeilijkheden zijn er van dit systeem van differentiaalvergelijkingen slechts benaderende oplossingen bekend. Bij met gas gevulde gloeilampen doet zich nog de bijkomstige moeilijkheid voor dat de symmetrie gering is en dat er extreme temperatuurverschillen voorkomen. Langmuir [1] heeft een benaderende oplossing voorgesteld. Hij vereenvoudigde het probleem tot een warmtegeleidingsprobleem, door aan te nemen dat de gloeidraad zou zijn omgeven door een stationaire gashuid. Van deze veronderstelling is later gebruik gemaakt door Elenbaas [2] voor de berekening van het massatransport. Hoewel de benadering van Langmuir tot een bevredigende beschrijving leidt van de totale massa- en warmteverliezen in normale met gas gevulde gloeilampen, is ze ongeschikt om er de scheikundige transportreacties in halogeenlampen mee te analyseren. Daartoe moet men de ruimtelijke concentraties van de verschillende wolfraam-halogeenverbindingen kennen, die afhankelijk zijn van de ruimtelijke temperatuur- en snelheidsverdeling van het edelgas. De analytische moeilijkheden maken het nodig deze ruimtelijke verdelingen experimenteel te bepalen. Gezien het verband tussen temperatuur- en snelheidsverdelingen, is het voldoende slechts één van deze grootheden te meten.

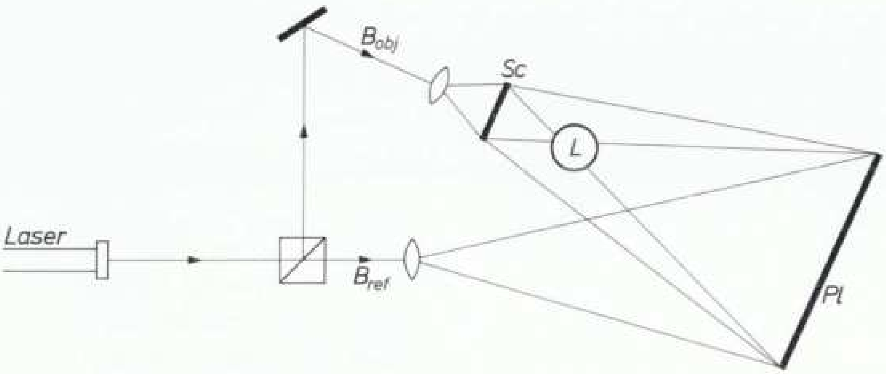


Fig. 1. De experimentele opstelling. Coherent licht van een laser wordt in twee bundels gesplitst, een objectbundel B_{obj} en een referentiebundel B_{ref}. Deze bundels komen weer samen en interfereren op de fotografische plaat *Pl*. Wordt de plaat ontwikkeld (het hologram) en laat men de referentiebundel erdoor vallen, dan verkrijgt men een golf die overeenkomt met de oorspronkelijke objectgolf. *L* is de te onderzoeken gloeilamp. *Sc* is een verstrooiingsplaat, waarmee een ruimtelijk uitgespreide lichtbron wordt verkregen.

Een fraaie methode ter bepaling van de temperatuurverdeling is de optische interferometrie. In de evenwichtstoestand is de druk in een brandende gloeilamp constant. Dit betekent dat de temperatuur T omgekeerd evenredig is met de dichtheid ϱ, die op haar beurt evenredig is met de brekingsindex n:

$$T^{-1} \sim \varrho \sim n.$$

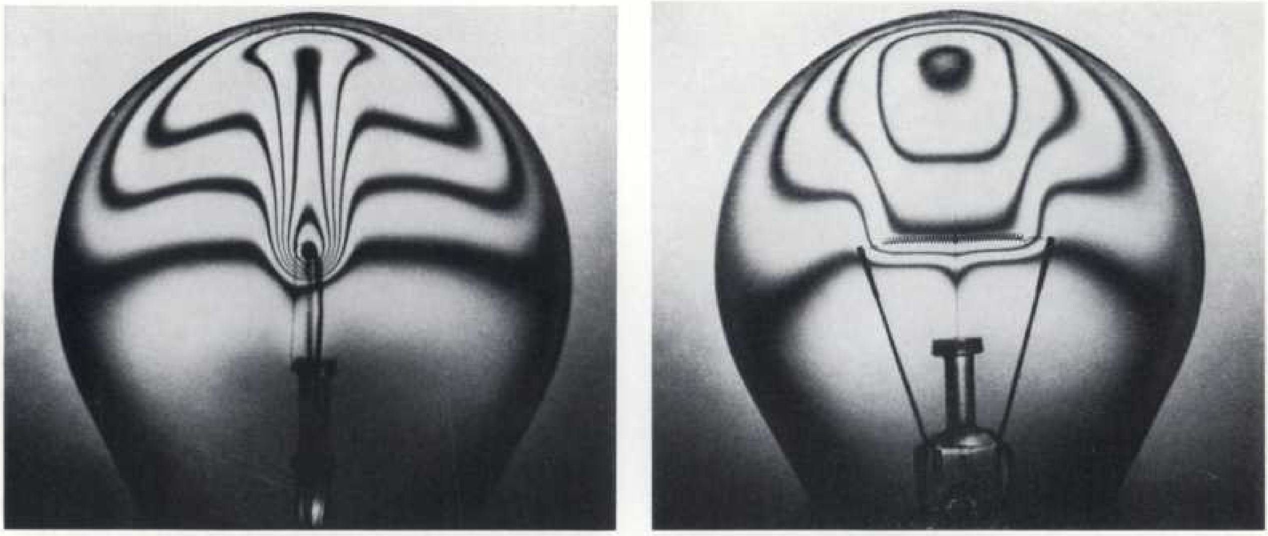
Slaagt men erin de verdeling van de brekingsindex $n(x,y,z)$ experimenteel te bepalen, dan kan men de temperatuurverdeling $T(x,y,z)$ dus berekenen.

Het meest gebruikte klassieke instrument voor het bepalen van veranderingen in de brekingsindex in gassen is de interferometer van Mach-Zehnder. Bij dit instrument leggen twee coherente lichtbundels verschillende optische wegen af, waarna ze worden samengebracht ten einde te interfereren. Uit de verandering van het interferentiepatroon die door het inbrengen van het te onderzoeken object wordt veroorzaakt, kan men dan de brekingsindex ervan afleiden. Bij gloeilampen heeft

Dr. H. Hörster en J. Fitzgerald, medewerkers van Philips Forschungslaboratorium Aachen GmbH, Aken, Duitsland.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023



a b

Fig. 2. Interferentiepatroon in een gloeilamp voor huishoudelijk gebruik (110 V, 100 W), verkregen door middel van interferentieholografie. Bij de holografische interferometrie met dubbele belichting worden twee opnamen gemaakt, bij twee verschillende toestanden van het object. Bij het reconstructieproces worden de twee respectieve objectgolven tegelijkertijd gereconstrueerd, zodat ze, als gevolg van hun coherentie, interfereren. In ons geval zijn de opnamen achtereenvolgens gemaakt met de lamp uitgeschakeld en brandend. *a)* Aanzicht in de richting van de gloeidraad. *b)* Zijaanzicht.

men evenwel te maken met *twee* brekingsindices, namelijk die van het glazen omhulsel en die van het edelgas, en alle veranderingen in het interferentiepatroon zijn een gevolg van hun gezamenlijk effect. Gezien de slechte optische eigenschappen van het glas, kan het effect ervan niet worden gecompenseerd door een identieke ballon in een andere lichtbundel te plaatsen. Het zou in principe mogelijk zijn, de wand van de ballon van optisch goede vensters te voorzien, maar daardoor zou de symmetrie gewijzigd worden en eveneens de temperatuurverdeling. Verdere nadelen van de interferometer zijn de hoge prijs en het feit dat men het apparaat nauwkeurig moet justeren.

Deze problemen kan men vermijden door holografische interferometrie toe te passen. In het algemeen wordt een hologram gevormd door het interferentiepatroon van twee coherente lichtbundels, een objectbundel en een referentiebundel, op een fotografische plaat vast te leggen. Het hologram — dat is de ontwikkelde plaat — heeft nu de eigenschap dat, als de referentiebundel erdoor gaat, een buigingsbeeld ontstaat dat met het object overeenkomt. Bij de holografische interferometrie met dubbele belichting wordt deze procedure twee keer uitgevoerd en worden golven, afkomstig van het object in twee verschillende toestanden, op dezelfde fotografische plaat vastgelegd. Deze beide golven worden gezamenlijk gereconstrueerd en interferen onderling, zodat een interferentiepatroon ontstaat.

In het geval van een gloeilamp wordt de eerste opname gemaakt terwijl de lamp uitgeschakeld is, de tweede terwijl de lamp brandt (*fig. 1*). Tijdens de eerste

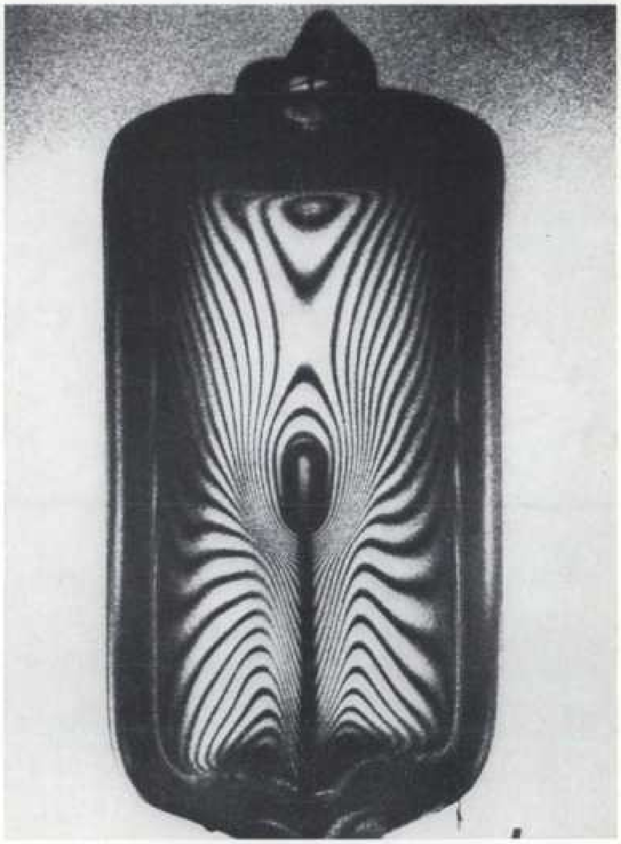


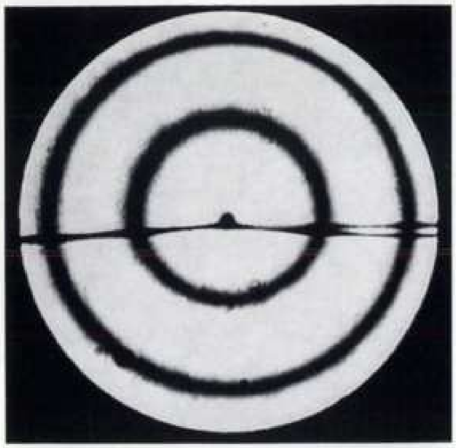
Fig. 3. Als fig. 2*a*, maar nu voor een halogeen-projectielamp (12 V, 100 W) gevuld met xenon. Druk 4 atm, lengte 17 mm, diameter 10 mm.

[1] I. Langmuir, Phys. Rev. **34**, 401, 1912.
[2] W. Elenbaas, Philips Res. Repts. **18**, 147, 1963.

Digitized in Heiloo the Netherlands Not for sale!

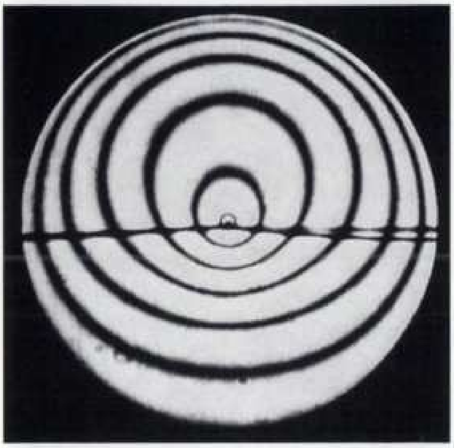
www.freeservicemanuals.info 11/14/2023

Fig. 4. Interferentiepatronen in een cilindersymmetrische, met argon gevulde lamp (diameter van de buisvormige ballon 4 cm; draaddiameter 200 μm), gezien in de richting van de gloeidraad. In deze opname stelt elke interferentielijn een isotherm voor. De invloed van drukveranderingen, bij gelijkblijvende temperatuur van de gloeidraad (1400 K), is duidelijk waarneembaar. In de tabellen is *m* de orde van de donkere interferentielijnen, geteld van het midden uit.



a) Druk 30 torr, diameter van Langmuir-laag 2,8 cm.

m	*T* (K)
1	1338
2	563
3	356



b) Druk 75 torr, diameter van Langmuir-laag 1,7 cm.

m	*T* (K)
1	946
2	680
3	532
4	436
5	370
6	313

c) Druk 200 torr, diameter van Langmuir-laag 1,0 cm.

m	*T* (K)
1	1295
2	1080
3	923
4	809
5	718
6	647
7	588
8	538
9	497
10	462
11	431
12	405
13	381
14	360
15	341

Fig. 5. Interferentiepatroon in een cilindersymmetrische lamp bij een gasdruk van 200 torr argon en een gloeidraadtemperatuur van 2350 K. Uit een vergelijking met fig. 4*c* blijkt dat bij de hogere temperatuur de temperatuurverdeling in de nabijheid van de gloeidraad beter met de symmetrische vorm overeenkomt.

m	*T* (K)	*m*	*T* (K)
1	2150	8	647
2	1615	9	588
3	1295	10	538
4	1080	11	497
5	923	12	462
6	809	13	431
7	718	14	405
		15	381
		16	360
		17	341

Druk 200 torr, diameter van Langmuir-laag 1,1 cm.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

opname heeft de dichtheid en dus ook de brekingsindex n_0 van het edelgas overal dezelfde waarde. Als de lamp brandt, is er een driedimensionele temperatuurverdeling en daarmee samenhangend een driedimensionele brekingsindexverdeling. Bij de reconstructie worden de twee opnamen gesuperponeerd; de beide gereconstrueerde objectgolven interferen dan, zodat een interferentiepatroon wordt verkregen. Daar de glazen ballon bij beide opnamen gelijk is, komt zijn aanwezigheid in het interferentiepatroon niet tot uiting.

Het verband tussen de orde m van de waargenomen interferentielijnen wordt gegeven door de integraal van de verandering in de brekingsindex langs de optische weg door de ballon:

$$m = \frac{1}{\lambda}\int [n(x,y,z) - n_0]\, \mathrm{d}s\,,$$

Hierin is λ de golflengte van het licht en ds een element van de weg. Voor een driedimensionele verdeling is de berekening van $n(x,y,z)$ met behulp van deze formule uiterst gecompliceerd. Onlangs hebben Junginger en Van Haeringen [3] echter aangetoond dat men door omkering van de integraal een analytische formule kan verkrijgen waarin $n(x,y,z)$ wordt uitgedrukt in de orde van de interferentielijnen.

Fig. 2 laat de interferentiepatronen zien van een gloeilamp voor huishoudelijk gebruik. Fig. 2*a* is het patroon, verkregen bij opname in de lengterichting van de gloeidraad, fig. 2*b* toont het patroon van de lamp met de gloeidraad in zijaanzicht. In deze laatste foto ziet men dat vlak bij de gloeidraad de orde over vrijwel de gehele lengte van de draad nagenoeg gelijk blijft. Laat men verschijnselen aan de uiteinden buiten beschouwing, dan mag men het patroon van interferentielijnen in fig. 2*a* in eerste benadering als een systeem van isothermen beschouwen. Als er inderdaad een stilstaande gashuid bestaat, zoals de theorie van Langmuir voorspelt, dan moet de temperatuurverdeling in de onmiddellijke nabijheid van de gloeidraad met de geometrische vorm van deze draad overeenstemmen. Dit blijkt volgens fig. 2*a* vlak onder de gloeidraad bij benadering wel het geval te zijn. De afwijking van de cilindersymmetrie boven de gloeidraad wordt veroorzaakt door sterke convectie. Dit is ook het geval bij de kleine halogeenlamp waarvan het interferentiepatroon in *fig. 3* is afgebeeld.

Om de moeilijkheden, verbonden aan de berekening van een driedimensionele brekingsindexverdeling te vermijden, hebben we de geldigheid van de theorie van Langmuir onderzocht in een cilindersymmetrische lamp. Deze was zo ontworpen dat de temperatuurverdeling werd weergegeven door het interferentiepatroon. In een voorafgaand onderzoek werd nagegaan wat de invloed van de gasdruk was op de temperatuurverdeling van het edelgas bij een constante gloeidraadtemperatuur van 1400 K. De resultaten zijn weergegeven in *fig. 4*. Men ziet dat bij een hogere gasdruk de invloed van de convectie groter is en de temperatuurverdeling een sterker afwijkend patroon vertoont. In de tabellen naast elk van de foto's zijn de temperatuur ter plaatse van elke interferentielijn en de dikte van de overeenkomstige Langmuir-laag vermeld.

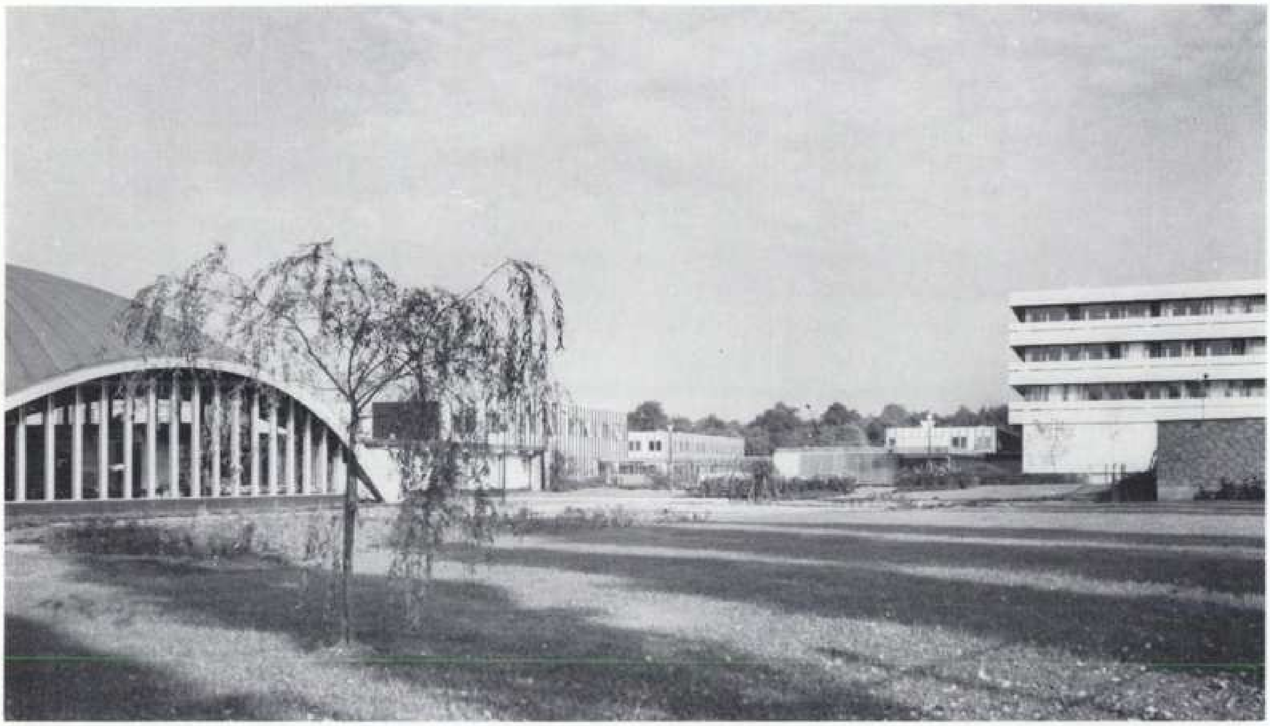
Vergelijkt men de temperatuurverdeling met de berekende dikte van de laag, dan blijkt de theorie van Langmuir uitsluitend te gelden bij een uitzonderlijk lage druk. Neemt de druk toe, dan doet de invloed van de convectie de cilindersymmetrische temperatuurverdeling verloren gaan. *Fig. 5* laat de temperatuurverdeling bij 200 torr nog eens zien, maar nu bij een veel hogere gloeidraadtemperatuur (2350 K). Hieruit blijkt dat de cilindersymmetrie in de onmiddellijke nabijheid van de draad bij de hogere temperatuur beter is, d.w.z. de afwijking van de theorie van Langmuir is geringer. Het opmerkelijke verschil tussen fig. 4*c* en fig. 5 moet worden toegeschreven aan het toenemen van de viscositeit van het gas met de temperatuur.

Dit voorlopige onderzoek toont aan dat het twijfelachtig is of de theorie van Langmuir wel zonder meer op gloeilampen mag worden toegepast. Een meer gedetailleerde behandeling van dit onderwerp zal worden gegeven in Philips Research Reports.

[3] H. G. Junginger en W. van Haeringen, te publiceren in Optics Communications.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023



Laboratoires d'Electronique et de Physique Appliquée, Limeil-Brévannes, Frankrijk.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Het tunneleffect in halfgeleider-metaal-contacten bij drukverhoging

P. Guétin en G. Schréder

Inleiding

We beschouwen het contact tussen een metaal en een sterk gedoteerde *N*-halfgeleider met vrije elektronen (*fig. 1*). Door het verschil in uittreearbeid tussen de beide media zal het bandenschema van de halfgeleider een duidelijke kromming vertonen in de omgeving van het contact, waar een ruimtelading verschijnt [1]. Er wordt een potentiaalbarrière gevormd die het de elektronen onmogelijk maakt van de ene zijde naar de andere te stromen, ook als er tussen beide materialen een spanning wordt aangelegd. De aanwezigheid van deze barrière is verantwoordelijk voor de gelijkrichting in de bekende Schottky-diodes. Dergelijke schakelingselementen worden gemaakt van materialen met een betrekkelijk geringe doteringsconcentratie n ($n \approx 10^{16}$ cm^{-3}), en de gelijkrichting hangt samen met het feit dat de elektronen de barrière moeten overwinnen om van de ene elektrode naar de andere te kunnen komen. Thermische excitatie is noodzakelijk om de elektronen *over* de barrière heen te helpen. Bij zwakke dotering en bij kamertemperatuur is deze „thermische emissie" het overheersende proces; dit geval is in fig. 1 schematisch aangeduid door de weg *1*. Bij een zelfde totale buiging van de band zal de breedte van het ruimteladingsgebied (d.i. de barrière) des te geringer zijn naarmate de concentratie van de dotering hoger is. Bij een zeer hoge concentratie van de dotering kan de barrière zo smal worden [1] (ca. 100 Å) dat er een niet te verwaarlozen waarschijnlijkheid is dat de elektronen het verboden deel van de band *passeren* dat de barrière vormt. Als de temperatuur zo laag is dat de thermische excitatie geen bijdrage levert, treedt alleen deze vorm van geleiding op, het zg. tunneleffect (weg *3*). Bij hogere temperaturen dragen zowel thermische excitatie als het tunneleffect bij tot het elektronentransport (weg *2*); dit geval duidt men in het algemeen aan met de term „thermische veldemissie" [2]. Het gaat ons hier in hoofdzaak om het zuivere tunneleffect bij lage temperatuur.

In feite is het tunneleffect een zeer algemeen verschijnsel in quantummechanische systemen. De emissie van α-deeltjes door atoomkernen is er een voorbeeld van. De eerste overtuigende waarneming van het effect in halfgeleiders werd in 1957 door Esaki gedaan [3]: de vondst van een *negatieve* differentiële weerstand in een sterk gedoteerde *P-N*-diode leverde het ondubbelzinnige bewijs voor het bestaan van een tunneleffect in het verboden deel van de energieband. Omstreeks 1960 nam het fundamentele onderzoek aan de zg. Esaki-diodes snel af en kreeg de ontwikkeling van toepassingen in elektronische apparaten en onderdelen de overhand. Enige tijd later gingen de fysici hun aandacht richten op de halfgeleider-metaal-contacten; het tunneleffect daarin werd ontdekt in 1966.

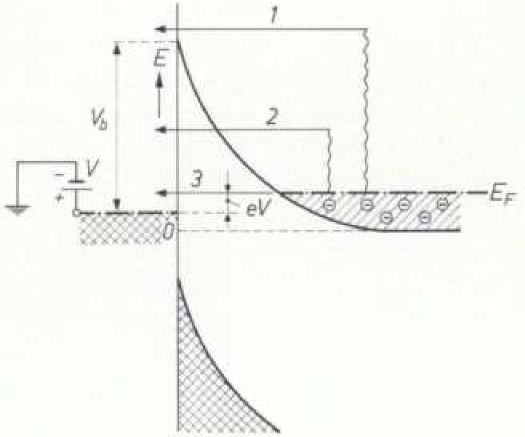


Fig. 1. Bandstructuur in de omgeving van het contact tussen een metaal en een *N*-halfgeleider. De totale energie *E* van de elektronen (de verticale as) wordt gerekend vanaf de onderkant van de geleidingsband. Langs de horizontale as is de afstand tot de contactplaats uitgezet. De halfgeleider is zo sterk gedoteerd dat alle toestanden in de geleidingsband tot aan het Fermi-niveau E_F door elektronen bezet zijn. Door de buiging van de banden ontstaat een potentiaalbarrière. *e* is de absolute waarde van de lading van een elektron, en *eV* is het energieverschil tussen de twee Fermi-niveaus, veroorzaakt door de aangelegde spanning *V*. (*V* is in deze figuur positief; de pluspool van de batterij is met het metaal verbonden.) Er kunnen drie geleidingsprocessen optreden, waarbij de elektronen over de barrière heen gaan door thermische excitatie, of er dwars doorheen: *1* thermische emissie, *2* „thermische veldemissie", *3* tunneleffect. In alle gevallen kan de weerstand van de materialen verwaarloosd worden ten opzichte van de overgangsweerstand, zodat de volle aangelegde spanning over het contact staat.

Dr. P. Guétin en dr. G. Schréder, medewerkers van Laboratoires d'Electronique et de Physique Appliquée, Limeil-Brévannes, Frankrijk.

[1] Zie b.v.: L. Heijne, Fysische grondslagen van de fotogeleiding, I. Grondbegrippen; contacten op halfgeleiders, Philips techn. T. 25, 262-273, 1963.
[2] F. A. Padovani, in: R. K. Willardson en A. C. Beer (redacteuren), Semiconductors and semimetals, Vol. 7A, Academic Press, New York 1971, hoofdstuk 2.
[3] L. Esaki, Phys. Rev. 109, 603, 1958.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Ofschoon dit type tunnelovergang niet erg geschikt bleek voor toepassingen in apparaten en onderdelen, heeft men wel snel onderkend dat deze overgangen zeer interessant fysisch onderzoek mogelijk maken. Dit komt voornamelijk door het „spectroscopische" karakter van het verschijnsel. Het tunneleffect geeft namelijk de mogelijkheid om vrije ladingdragers te injecteren waarvan de energie op een *willekeurig niveau* boven het Fermi-niveau ligt. Aangezien alle elektronen aan de Fermi-Dirac-statistiek voldoen, ontstaat de stroom doordat er elektronen van een bezette toestand aan de ene kant van de barrière overgaan naar een lege toestand aan de andere kant. Zonder dat we de zaken verder uitdiepen, zal het duidelijk zijn dat een zorgvuldig onderzoek van het geleidingsvermogen belangrijke gegevens verschaft over de verdeling der beschikbare toestanden, dat is hun dichtheid en bezetting. Met andere woorden: het tunneleffect vormt een gevoelige analysator van de bandstructuur. In vaste stoffen bestaan echter geen werkelijk vrije ladingdragers: in feite zijn de ladingdragers daar doorlopend in wisselwerking met hun omgeving, dat wil zeggen: met andere elektronen, atomen van het rooster, stoorcentra, enz. Tot op zekere hoogte vindt men deze wisselwerkingen weerspiegeld in het geleidingsvermogen; zij geven aanleiding tot kleine anomalieën bij bepaalde waarden van de aangelegde spanning. Het tunneleffect kan derhalve gebruikt worden als een middel voor het bestuderen van deze wisselwerkingen der elektronen.

Het doel van onze eigen experimenten was voornamelijk nieuw licht te laten schijnen op twee van deze aspecten: de samenhang tussen de bandstructuur en de geleiding door middel van het tunneleffect, en de wisselwerking van de elektronen met roostertrillingen (fononen).

Ofschoon verschillende soorten tunnelovergangen voor ons doel bruikbaar zijn, hebben we om de volgende redenen halfgeleider-metaal-contacten gekozen. In de eerste plaats kan men ze veel reproduceerbaarder en met veel geringere verontreiniging maken dan andere overgangen. Verder kunnen de belangrijkste parameters van deze contacten door onafhankelijke proeven bepaald worden, zodat men mag hopen op een kwantitatieve overeenstemming tussen theorie en experiment. Ten slotte zijn halfgeleider-metaal-contacten goed bestand tegen grote wisselingen van temperatuur en druk. De meeste van deze kwaliteiten hangen samen met het feit dat de barrière geen kunstmatige laag is, maar voortkomt uit de „intrinsieke" eigenschappen van het systeem halfgeleider/metaal.

Nadat we enig kwantitatief inzicht hadden verworven in de geleiding door het tunneleffect in eenvoudige isotrope materialen, zijn we aandacht gaan besteden aan de invloed van de bandstructuur op de totale geleiding en op de wisselwerking tussen elektronen en fononen. Men mag in het algemeen stellen dat de voorstelling van de geleidingsband van halfgeleiders in de zg. ***k***-ruimte (***k*** is de golfvector van een elektron) vele potentiaalkuilen („dalen") vertoont in de buurt van punten met sterke kristalsymmetrie [4]. Zo wordt het Gunn-effect in GaAs veroorzaakt doordat er elektronen onder invloed van sterke elektrische velden worden verplaatst naar potentiaaldalen met een hogere energie, waar ze totaal andere transporteigenschappen hebben.

Door druk op een halfgeleider uit te oefenen kan men op een elegante en goed beheersbare wijze de bandstructuur en in het bijzonder het energieverschil tussen de diverse subbanden veranderen. We hebben deze werkwijze samen met het tunneleffect gebruikt bij halfgeleiders met een „directe" (d.w.z. bij één zelfde waarde van ***k*** optredende) minimale bandafstand [5], zoals GaAs, en ook bij halfgeleiders met een „indirect" minimum, zoals Ge. In een bepaald materiaal (GaSb) is het ons gelukt met behulp van tunneleffectmetingen waar te nemen hoe bij drukverandering de directe minimale bandafstand geleidelijk toeneemt en ten slotte omslaat in een indirecte minimale bandafstand [6].

Dergelijke experimenten zijn van veel belang voor halfgeleiders waarin het Gunn-effect kan optreden. Het energieverschil tussen naburige potentiaalkuilen, dat door middel van het tunneleffect bij drukverhoging rechtstreeks kan worden gemeten, is hierbij namelijk een fundamentele parameter.

We zullen eerst het karakteristieke gedrag van een *N*-GaAs/metaal-contact bespreken, waarbij we zullen laten zien hoe de betrekkelijk geringe verandering van het geleidingsvermogen bij drukverhoging dit soort contacten bijzonder geschikt maakt voor een gedetailleerd onderzoek van de supergeleiding die het betrokken metaal bij drukverhoging vertoont. Daarna bespreken we de bijzonder grote veranderingen die in *N*-GaSb/metaal-contacten bij drukverhoging optreden en die gepaard gaan met de „omslag" van de minimale bandafstand, en laten we zien welke gegevens men uit de tunneleffectkarakteristieken kan verkrijgen. Alvorens we echter onze metingen bespreken, willen we van de gebruikte materialen de bandstructuur in de ***k***-ruimte nader specificeren.

[4] Voor een uitvoeriger uiteenzetting zie b.v.: J. S. Blakemore, Solid state physics, uitg. Saunders, Philiadelphia 1970.

[5] P. Guétin en G. Schréder, Solid State Comm. **9**, 591, 1971 (no. 9). „Direct" en „indirect" impliceren dat de overgang van een elektron uit de valentieband naar de geleidingsband bij minimaal energieverschil mogelijk is door de absorptie van één foton (een verticale overgang in de ***k***-ruimte, dus bij constante ***k***) resp. door één foton in samenwerking met één fonon (een niet-verticale overgang, waarbij dus twee verschillende waarden van ***k*** zijn betrokken).

[6] P. Guétin en G. Schréder, Phys. Rev. Letters **27**, 326, 1971.

[7] De overdracht van elektronen die leidt tot het Gunn-effect gebeurt bij een sterker elektrisch veld en een hogere temperatuur dan gebruikt worden bij de tunnel-experimenten.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Beïnvloeding van de bandstructuur door hydrostatische druk

We zullen hier voor de twee verbindingen die onze meeste aandacht hebben gehad, GaAs en GaSb, de verandering beschrijven die de geleidingsband ten opzichte van de valentieband ondergaat bij verandering van de hydrostatische druk. *Fig. 2* toont van elk van deze stoffen voor een bepaalde kristalrichting een schematische voorstelling van de dispersierelatie $E(k)$, dat is de energie der elektronen als functie van de golfvector $\boldsymbol{k}$ waarmee, naast een impulsvector $\boldsymbol{p}$ ($= \hbar \boldsymbol{k}$), hun quantummechanische toestandsfunctie geassocieerd is. Van deze curven zijn alleen de gedeelten afgebeeld in de buurt van de twee extrema die de minimale afstand tussen geleidingsband en valentieband bepalen. Met andere woorden: we beschouwen alleen het diepste „dal" en de hoogste „top" van de geleidings-, resp. de valentieband. De getrokken lijnen geven de configuratie bij atmosferische druk weer, de onderbroken lijnen die bij hoge druk.

We zien in fig. 2 dat bij GaAs de minimale energie van het Γ-dal dat symmetrisch ten opzichte van $k = 0$ ligt, bij drukverhoging stijgt, terwijl de energie van het daarnaast gelegen X-dal juist daalt. Bij een druk van 17 kbar, de hoogste druk die we in onze experimenten hebben gebruikt, blijkt het energieverschil tussen deze dalen verminderd te zijn van 0,36 eV (bij atmosferische druk) tot ca. 0,16 eV. Dit is echter nog niet genoeg om bij zeer lage temperaturen een meetbaar transport van elektronen uit het Γ-dal naar de X-dalen mogelijk te maken. De elektronen blijven in het „centrale" dal en houden bijgevolg hun karakteristieke eigenschappen, ongeacht de druk: we kunnen dus geen sterke bandstructuureffecten verwachten.

GaSb vertoont bij drukverhoging een interessanter gedrag dan GaAs. De toestanden met minimale energie zijn bij deze verbinding gelokaliseerd in het Γ-dal dat symmetrisch om $k = 0$ ligt, en in vier dalen in de (111)-richtingen, die de vorm van ellipsoïden hebben. Bij atmosferische druk is de minimale bandafstand direct en bedraagt het energieverschil $\Delta E = E_L - E_\Gamma$ tussen het naastgelegen dal, het L-dal, en het centrale dal ca. 0,1 eV. Wanneer de druk nu wordt verhoogd, bewegen de beide minima zich naar hogere waarden; het bijzondere is hierbij dat het minimum van het Γ-dal dit veel sneller doet dan het minimum van het L-dal, zodat het dit al spoedig inhaalt. Boven 10 kbar blijkt het Γ-dal een energieminimum te hebben dat hoger ligt dan het minimum van het L-dal: de minimale bandafstand is dan niet langer direct, maar is indirect geworden. De omslag van de minimale bandafstand gebeurt in de buurt van de genoemde druk van 10 kbar. Tijdens dit proces worden steeds meer elektronen uit het Γ-dal naar het L-dal overgebracht, waar ze heel andere eigenschappen hebben. Zo is de effectieve massa in de toestanden van het L-dal groter dan in die van het Γ-dal: $m_L^* \approx 2\, m_\Gamma^*$.

We zien dus dat, ofschoon de bandstructuur van GaSb bij lage druk lijkt op die van GaAs, deze structuur bij hoge druk een grote overeenkomst vertoont met die van Ge. Metingen van het tunneleffect bevestigen niet alleen deze analogie, maar ze stellen ons ook in staat om zeer nauwkeurig de veranderingen waar te nemen die leiden tot de omslag van de minimale bandafstand.

Nadat we aldus het kader hebben aangegeven waarbinnen ons onderzoek aan het tunneleffect is uitgevoerd, zullen we nu overgaan tot het bespreken van de experimenten.

De uitvoering van de experimenten

De halfgeleider-metaal-contacten hebben we gemaakt door in ultrahoog vacuüm een staafje halfgeleider te kloven terwijl het gewenste metaal er werd opgedampt. Onmiddellijk na het splijten werd voor het schone oppervlak een masker met gaatjes geplaatst; gedurende het opdampen hielden we de druk in het va-

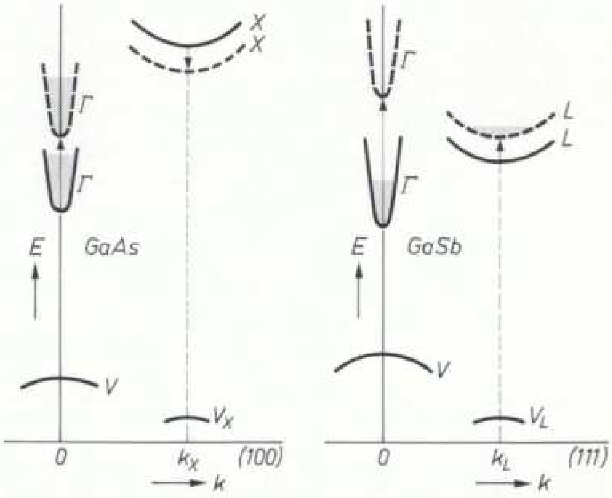


Fig. 2. Schematische voorstelling van de energiebanden van GaAs en GaSb in de $\boldsymbol{k}$-ruimte, resp. voor de kristalrichtingen (100) en (111). De curven aangeduid met V, V_X en V_L geven de gebieden rond de maxima van de valentieband weer; Γ, X en L zijn de dalen waarin de minima van de geleidingsbandenergie optreden. Getrokken lijnen, resp. streeplijnen hebben betrekking op atmosferische druk, resp. sterke overdruk. In beide stoffen daalt bij drukverhoging het energieverschil tussen de twee soorten dalen van de geleidingsband. De grootste bij onze experimenten beschikbare drukverhoging is niet groot genoeg om bij GaAs een waarneembaar transport van elektronen vanuit de Γ-minima naar de X-minima mogelijk te maken [7]. In GaSb kunnen daarentegen de minima van de L-dalen worden verschoven naar een energie lager dan die van het tegelijk omhoog schuivende Γ-dal. Bij een grote drukverhoging wordt derhalve de kleinste bandafstand indirect. De omslag van direct naar indirect gebeurt bij $P = P_0$, de druk waarbij de minima van het Γ-dal en de L-dalen samenvallen.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

cuümsysteem beneden 7×10^{-7} Pa (d.i. 5×10^{-9} torr). Na het opdampen werden de elektrische aansluitingen aan de metaal-„stippen" aangebracht door er met een druppeltje zilverpasta een dun draadje op te plakken. Het Ohm-contact aan de kant van de halfgeleider was al met een van de klassieke legeringsmethoden aangebracht voordat het materiaal in het vacuüm kwam.

De opstelling bestond in hoofdzaak uit een hogedrukvat, gevuld met isopentaan (2-methylbutaan) als drukoverbrengende vloeistof. De preparaten bevonden zich hierin op een afdichtstop met zeven elektrische doorvoeren. Het vat stond via een capillair in verbinding met het hoofddrukvat, waarin de druk met een manganinemanometer werd gemeten. Bij het afkoelen van het vat bevriest het isopentaan bij een temperatuur die afhangt van de gekozen druk (bij 3 kbar is die temperatuur ongeveer 145 K en bij 16 kbar 260 K). De druk werd gedurende het experiment constant gehouden.

Voor het bepalen van de karakteristieke grootheden van het tunneleffect hebben we een conventionele schakeling gebruikt, die een gelijkstroom levert waarop een kleine, constante wisselstroom met een frequentie van 500 Hz gesuperponeerd is. Het spanningssignaal met de grondfrequentie, dat evenredig is met de differentiële weerstand dV/dI, werd verwerkt in een fasedetectieschakeling („lock-in") en het uitgangssignaal daarvan werd automatisch uitgezet als functie van de aangelegde spanning. De tweede harmonische werd ook gedetecteerd; dit signaal is een directe maat voor d^2I/dV^2, dus voor de afgeleide van het differentiële geleidingsvermogen naar de aangelegde spanning. De hoofdpunten van het tunneleffect zijn kwalitatief eenvoudig te verklaren: de breedte van de potentiaalbarrière moet voor de snelste elektronen die „tunnelen", zijn grootste waarde hebben bij een aangelegde spanning nul. Zowel bij positieve als bij negatieve spanning wordt de breedte kleiner, waardoor de ermee samenhangende tunnelkans toeneemt. Dit betekent dus dat de differentiële weerstand daalt als de aangelegde spanning toeneemt, onverschillig welk teken deze heeft. De gemeten weerstand vertoont zijn maximale waarde inderdaad bij een spanning die nagenoeg nul is (bij $V \approx 50$ mV).

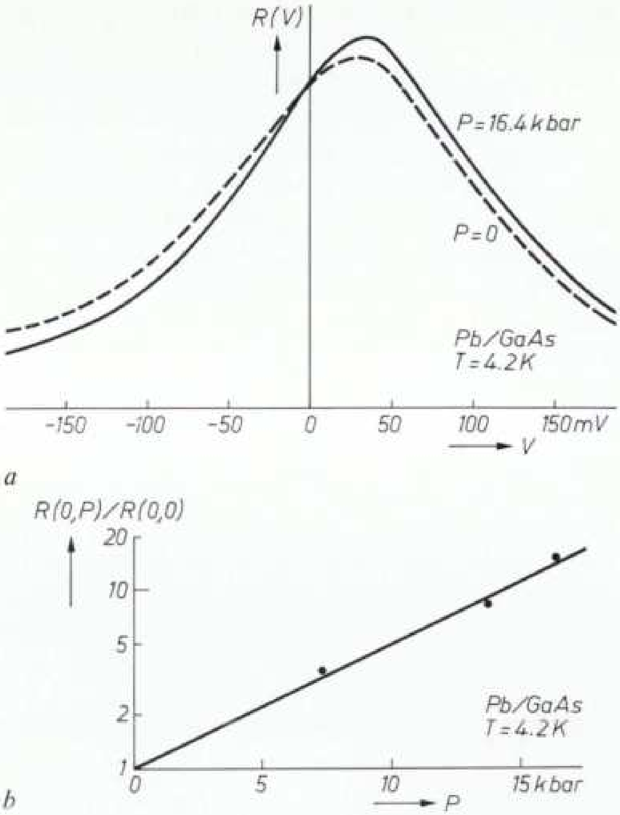


Fig. 3. *a*) De differentiële weerstand $R(V)$, d.i. dV/dI, van een N-GaAs/lood-contact bij een temperatuur van 4,2 K, uitgezet tegen de aangelegde spanning V bij atmosferische druk en bij 16,4 kbar. De doteringsconcentratie bedraagt $5{,}4 \times 10^{18}$ cm^{-3}. Gemakshalve zijn beide curven genormeerd op $V = 0$. *b*) Deze halflogaritmische grafiek illustreert de drukafhankelijkheid van $R(0)$, de weerstand bij spanning nul, voor een N-GaAs/lood-contact.

De invloed van druk op N-GaAs/metaalcontacten

Zoals reeds is besproken, mag men voor een N-GaAs/metaal-contact bij drukverhoging geen ingrijpende verandering van de bandstructuur in het GaAs verwachten, maar slechts een zekere toeneming van de bandafstand. Deze toeneming betekent een vergroting van de hoogte van de potentiaalbarrière en van de effectieve massa. Anderzijds mogen we de drukafhankelijkheid van de diëlektrische constante, van de dichtheid der ladingdragers, en van de oppervlakte van de grenslaag verwaarlozen.

De differentiële weerstand van een dergelijk contact is uitgezet in *fig. 3a*, zowel bij atmosferische druk als bij 16,4 kbar. Door normeren zijn de twee waarden bij $V = 0$ aan elkaar gelijk gemaakt, we zien nu duidelijk dat de curve maar weinig is veranderd. In fig. 3*b* is de relatieve verandering van de differentiële weerstand bij $V = 0$ uitgezet als functie van de drukverhoging. Men ziet dat er een (exponentiële) stijging optreedt, met een factor 10 per ongeveer 15 kbar. Bij het tunneleffect moeten dergelijke stijgingen als klein worden beschouwd, ze hangen hoofdzakelijk samen met de toeneming van de bandafstand en met de daardoor veroorzaakte verhoging van de barrière V_b.

Doordat in GaAs bij drukverhoging slechts betrekkelijk kleine veranderingen optreden, zijn de contacten van dit halfgeleidermateriaal met een metaal bijzonder geschikt voor de bestudering van de eigenschappen van het metaal bij drukverhoging. Een voorbeeld hiervan is het onderzoek aan supergeleidend lood, gedaan met behulp van N-GaAs/lood-contacten. In de superge-

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

leidende toestand [8] treedt via roostertrillingen een sterke wisselwerking tussen paren elektronen op. In afwijking van de normale toestand vindt men bij supergeleiding dat het energiespectrum der elektronen een verboden zone bevat die zich uitstrekt van $E_F - \Delta$ tot $E_F + \Delta$, waarbij E_F het Fermi-niveau in het metaal is. Bij het tunnelen gaat het altijd om elektronen die vanuit een bezette toestand overgaan naar een nog vrije toestand; het zal dus duidelijk zijn dat er bij temperaturen die nagenoeg samenvallen met het absolute nulpunt, geen stroom door het contact kan vloeien zolang de absolute waarde van de aangelegde spanning kleiner is dan Δ, de halve breedte van de verboden zone bij supergeleiding. De inzet van *fig. 4* illustreert deze bijzondere eigenschap van de tunnelstroom bij 4,2 K; tevens blijkt hieruit dat op deze wijze de breedte van de verboden zone rechtstreeks gemeten kan worden. Voor lood zijn de resultaten van dergelijke metingen in fig. 4 weergegeven. De breedte 2Δ van de zone neemt bij drukverhoging lineair af, en wel van 2,7 meV bij atmosferische druk tot 2,3 meV bij 15 kbar. Deze resultaten zijn goed vergelijkbaar met de gegevens die voor lagere drukken bij metaal-oxide-metaal-overgangen (M-O-M) zijn gevonden. In deze M-O-M-structuren bestaat de potentiaalbarrière uit een ongeveer 30 Å dikke oxidelaag, die in het geheel geen gaatjes, hoe klein ook, mag bevatten.

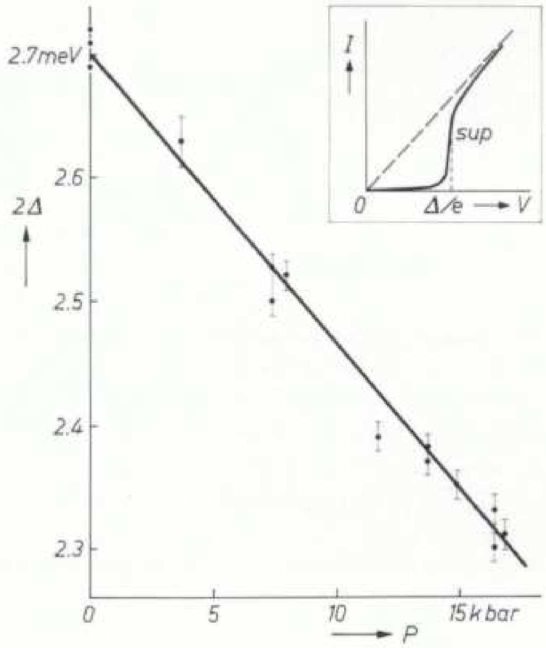


Fig. 4. De breedte 2Δ van de verboden zone bij supergeleidend lood, uitgezet als functie van de druk. De inzet toont de I,V-karakteristieken van het tunnelcontact. De streeplijn geeft de gewone geleiding in het metaal weer, de getrokken curve de supergeleiding. Zolang de aangelegde spanning beneden de waarde Δ/e blijft, zijn er voor elektronen van de halfgeleider geen vrije toestanden in het metaal bereikbaar en is de stroom dus nagenoeg nul. Boven $V = \Delta/e$ stijgt de stroomsterkte snel en bereikt al spoedig de normale waarde.

Het is hierdoor zeer moeilijk om dergelijke overgangen aan hoge drukken bloot te stellen zonder deze uiterst dunne laag te beschadigen; bij drukken boven 10 kbar is dit zelfs nagenoeg onmogelijk. Halfgeleider-metaalcontacten blijken in dit opzicht veel sterker te zijn dan de M-O-M-overgangen; zij vormen dan ook momenteel het beste preparaat voor het tunnelonderzoek.

De metingen kunnen worden uitgevoerd aan alle mogelijke supergeleidende metalen en legeringen. Men mag in het algemeen stellen dat metalen die een drukafhankelijke onregelmatigheid vertonen van de toestandsdichtheid in de buurt van het Fermi-niveau, goed onderzocht kunnen worden met behulp van een halfgeleider-metaal-contact, mits de halfgeleider, net als GaAs, in het desbetreffende drukgebied slechts geleidelijke en kleine veranderingen vertoont. We denken hier in het bijzonder aan het contact van een halfmetaal en een halfgeleider, waar het ontstaan van zo'n verboden zone door supergeleiding zichtbaar zou moeten worden als een weerstandspiek bij $V = 0$. Wat de halfgeleiders betreft, bestaat er vanzelfsprekend wel belangstelling voor de drukafhankelijkheid van de energieën der fononen; deze kan worden afgeleid uit de verschuivingen die de curven van de tweede afgeleide (d^2I/dV^2) vertonen.

Nadat we hebben laten zien hoe met deze techniek metalen onderzocht kunnen worden, zullen we nu het onderzoek van de bandstructuureffecten in halfgeleiders behandelen.

De invloed van druk op *N*-GaSb/metaal-contacten

In dit hoofdstuk zullen we in grote lijnen de belangrijkste verschillen bespreken tussen het tunnelen bij configuraties met een directe en bij die met een indirecte minimale bandafstand. We zullen hierbij veel aandacht besteden aan de verschijnselen die zich afspelen bij het omslaan van de minimale bandafstand. Eerst volgt nu een overzicht van de belangrijkste eigenschappen van de differentiële weerstand bij verschillende drukken; we zullen laten zien hoe hieraan belangrijke gegevens over de bandstructuur kunnen worden ontleend.

In *fig. 5* is de weerstand $R(0)$ van een *N*-GaSb/lood-contact bij een aangelegde spanning $V = 0$ uitgezet als functie van de drukverhoging. De metingen zijn uitgevoerd aan een representatief preparaat met een doteringsconcentratie van $3{,}5 \times 10^{18}$ cm^{-3}. De punten stellen de experimentele gegevens voor; de curve geeft het theoretisch berekende verloop weer. Vergelijken we deze figuur met fig. 3*b*, dan zien we dat de gevoeligheid voor drukverhoging van GaSb ruwweg 10^4 maal zo groot is als die van GaAs. Een ander opvallend ver-

[8] J. Volger, Philips techn. T. **29**, 1, 1968.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

schijnsel, dat bij de N-GaAs/metaal-contacten in het geheel niet voorkomt, is de knik in de kromme; $R(0)$ springt daar plotseling met meer dan twee ordes van grootte omhoog. De druk P_1 (12 kbar) waarbij deze discontinuïteit optreedt, is iets groter dan de verwachte „inversie"-druk P_0. Beide verschijnselen worden veroorzaakt doordat de minimale bandafstand bij drukverhoging van direct naar indirect omslaat, waarbij elektronen in de halfgeleider overgaan van het Γ-dal naar de L-dalen.

Het optreden van de discontinuïteit in de weerstand wordt duidelijk verklaarbaar als we bedenken dat de transmissiecoëfficiënt in benaderde vorm luidt:

$$D \approx \exp\left[-2\int_{x_1}^{x_2} \mathrm{k}(x)\mathrm{d}x\right],$$

met

$$\mathrm{k}(x) \approx \left[\frac{2m^*}{h^2}\{V(x)-\omega+\varepsilon_{11}\}\right]^{\frac{1}{2}} \approx v(m^*)^{\frac{1}{2}}\,.$$

Hoe kleiner de effectieve massa is, des te groter is de transmissiecoëfficiënt. Beneden 10 kbar wordt het tunneleffect hoofdzakelijk veroorzaakt door de elektronen die gelokaliseerd zijn in het Γ-dal en die een kleine effectieve massa hebben ($m_\Gamma^* \approx 0{,}05\ m_0$); deze kunnen de barrière betrekkelijk gemakkelijk passeren. Boven 12 kbar is de minimale bandafstand indirect geworden. In de buurt van $V = 0$ komen de tunnelende elektronen dan uit de L-dalen; zij worden gekarakteriseerd door een grote effectieve massa in de tunnelrichting ($m_L^* \gtrsim 0{,}1\ m_0$), wat maar een geringe transmissie betekent.

Aangezien het argument van de exponentiële functie waarden heeft die normaal tussen 15 en 20 liggen, zal het betrekkelijk kleine verschil in effectieve massa toch leiden tot een grote verandering van de weerstand.

Twee punten zullen we meer in detail bespreken: de overeenstemming tussen experiment en theorie voor de beide configuraties, en de ligging van de discontinuïteit in $R(0)$.

Bij de configuratie met de directe kleinste bandafstand stemmen de berekende en de waargenomen weerstandswaarden zeer goed overeen. Hieruit kunnen we concluderen dat de stroom, afkomstig uit het Γ-dal, de rechtstreekse tunnelstroom is (dus weg *3* in fig. 1). Bij de indirecte configuratie is de waargenomen weerstand ruwweg 75 maal zo klein als de berekende, wat suggereert dat er parallel met de gebruikelijke rechtstreekse weg nog een tweede weg bestaat. Deze discrepantie kan worden toegeschreven aan diffusie van elektronen naar geïoniseerde stoorcentra; in dit artikel zullen we daar niet verder op ingaan.

Aangezien de effectieve massa in het centrale dal het kleinst is, zal het grootste deel van de stroom daaruit afkomstig zijn, althans zolang er daar bezette toestanden zijn. De steile sprong treedt op bij de druk waarbij alle elektronen uit het Γ-dal naar de L-dalen zijn overgebracht, zodat het Γ-dal geheel leeg is. De druk P_1 waarbij deze situatie optreedt, is slechts weinig groter dan P_0 : $P_1 \approx P_0 + 1{,}5$ kbar.

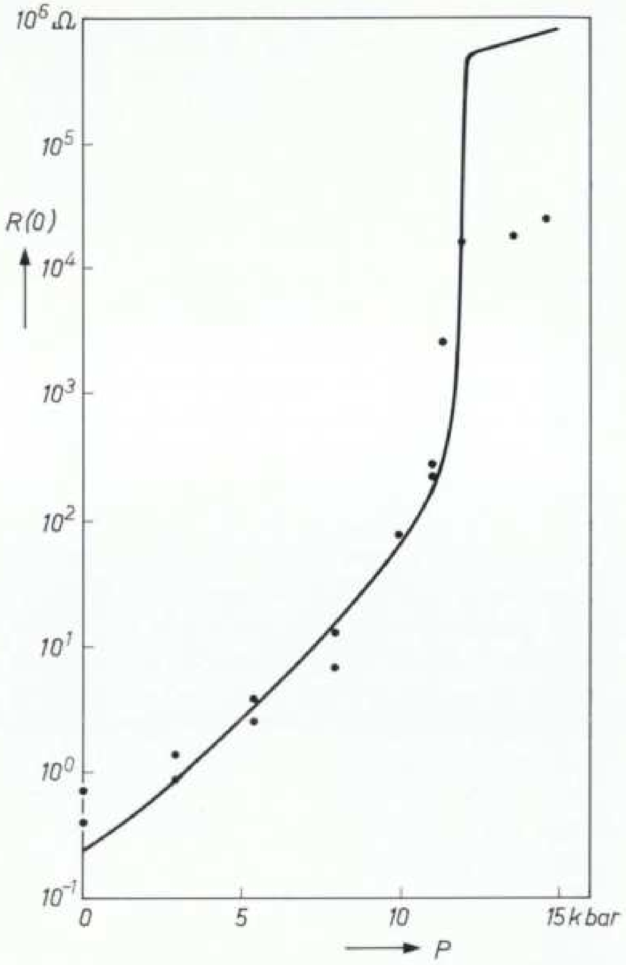


Fig. 5. De differentiële weerstand van een N-GaSb/lood-contact bij $V = 0$, uitgezet tegen de druk. De punten hebben betrekking op experimentele gegevens en de getrokken curve is gevonden door numerieke berekeningen. De steile sprong in de weerstand doet zich voor bij $P = P_1$, dat is de druk waarbij het Γ-dal volledig onbezet is geraakt. Deze discontinuïteit ontstaat doordat de effectieve massa van de betrokken tunnelende elektronen verandert. Voor $P \lesssim 10$ kbar is de kleinste bandafstand direct en voor $P > P_1$ (ca. 12 kbar) is die afstand indirect geworden.

Nu we enig inzicht hebben verworven in de absolute waarde van de differentiële weerstand bij aangelegde spanning nul, kunnen we de spectroscopische kant ervan, met andere woorden het verloop van deze weerstand als functie van de spanning V, nader gaan bekijken. *Fig. 6* toont vier curven die representatief zijn voor bepaalde configuraties: de onderste twee horen bij de „directe" configuratie, de bovenste twee bij de „indirecte". De streeplijnen geven de resultaten weer van berekeningen; deze zijn gemakshalve door normeren bij $V = 0$ aangepast aan de experimentele waarde. Opnieuw zien we hier dat de curven opgenomen bij de laagste twee drukken goed met de theoretische lijnen overeenstemmen; de beide andere vertonen twee belangrijke verschijnselen die toegeschreven kunnen worden aan de bandstructuur van GaSb. Ten eerste ver-

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

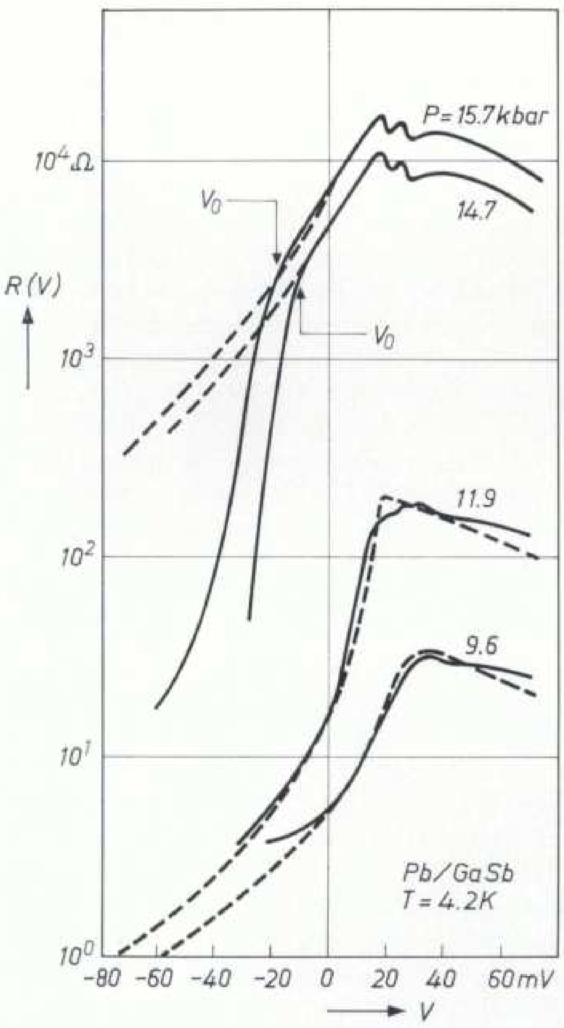


Fig. 6. De differentiële weerstand $R(V)$ van een N-GaSb/lood-contact uitgezet tegen de aangelegde spanning bij vier verschillende drukken. De streeplijnen geven de berekende curven, die door normeren bij $V = 0$ aangepast zijn aan de experimentele curven. De beide onderste curven (P = 9,6 en 11,9 kbar) zijn representatief voor drukken kleiner dan P_1; de weerstand wordt dan voornamelijk bepaald door het Γ-dal. De vorm van deze curven stemt goed overeen met de theoretische verwachting.

De curven bij de hoge drukken (P = 14,7 en 15,7 kbar) zijn typerend voor drukken groter dan P_1, waarbij de kleinste bandafstand indirect is. In de buurt van $V = 0$ stemmen de experimentele gegevens zeer goed overeen met een berekening waarin alleen de L-dalen beschouwd worden.

Deze curven vertonen twee opvallende verschijnselen. De differentiële weerstand daalt steil bij spanningen die sterker negatief zijn dan een bepaalde, drukafhankelijke, drempel V_0 — de „knik”. Bij positieve spanning treden twee sprongen in de weerstandscurve op, en wel bij spanningen die uiterst weinig drukafhankelijk zijn. Uit beide verschijnselen blijkt dat er naast het gewone, rechtstreekse geleidingsproces vanuit het L-dal, nog andere processen tot de geleiding kunnen bijdragen.

tonen de gemeten curven een benedenwaartse knik bij spanningen die lager zijn (meer negatief) dan V_0 (zie pijl), waarbij de drempel V_0 van de druk afhangt. Ten tweede vertonen de curven twee duidelijke sprongsgewijze dalingen van de weerstand bij spanningen die nauwelijks van de druk afhangen. Deze beide verschijnselen zullen we nu achtereenvolgens bespreken.

De knik in de weerstandscurve bij negatieve spanning

Bij drukken hoger dan P_1 zijn de L-dalen de diepste dalen. De gestreepte gedeelten van de bovenste twee curven van fig. 6 geven het berekende rechtstreekse tunneleffect naar deze L-dalen weer. Tussen $V = 0$ en $V = V_0$ is de overeenstemming met de experimenten bevredigend. De sterke en plotselinge daling van de weerstand voorbij de V_0-drempel wijst echter duidelijk op het in werking treden van een extra geleidingsproces, dat het rechtstreekse tunnelen naar de L-toestanden versterkt. Hoe hoger de druk is, des te sterker negatief is de waarde V_0 van de drempel. De knik in de weerstand kunnen we verklaren uit de bandstructuur in de buurt van het contact (*fig. 7*). Als $P > P_1$ is, ligt het Γ-dal bij een hogere energie dan de L-dalen en zijn alle toestanden in dit dal onbezet. Bij een spanning van ongeveer nul is de geleiding afkomstig van de stroom I_1 naar de L-toestanden, de enige toestanden die bij een dergelijke energie beschikbaar zijn. Bij spanningen die sterker negatief zijn dan een bepaalde drempel V_0, komen lege Γ-toestanden ter beschikking met een energie ter grootte van het Fermi-niveau in het metaal. Elektronen kunnen dan rechtstreeks naar deze Γ-toestanden overgaan. Dit resulteert in een stroom I_2, die I_1 versterkt. De hierbij optredende barrière (de streeplijn in fig. 7) is iets hoger en breder dan die voor de L-toestanden. Deze parameters blijken zodanige waarden te hebben dat $I_2 \gg I_1$ is; hiermee hebben we de oorzaak gevonden van de snelle afneming van de weerstand voor $V < V_0$. Dit verschijnsel verschaft ons een zeer rechtstreekse methode voor het meten van ΔE, d.i. het energieverschil tussen het L- en het Γ-dal. Hiervoor geldt namelijk de betrekking:

$$eV_0(P) = \Delta E(P) - E_F^{(L)}.$$

Het ontaarde Fermi-energieniveau in de L-dalen, $E_F^{(L)}$,

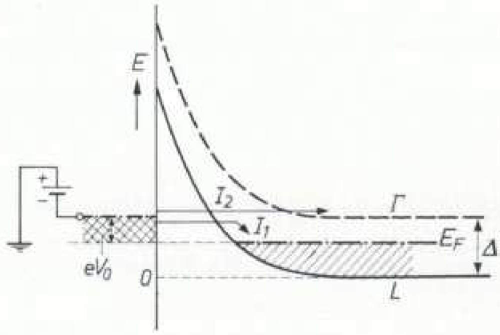


Fig. 7. Geschematiseerde bandstructuur in het gebied van het contact bij $P > P_1$. De minimale energie van het Γ-dal ligt boven die van het L-dal; de Γ-toestanden zijn onbezet. Bij spanningen die sterker negatief zijn dan de drempel V_0, kunnen elektronen rechtstreeks in het Γ-dal geïnjecteerd worden, wat gepaard gaat met een sterke weerstandsdaling vanwege de veel kleinere effectieve massa ($m_\Gamma^* \ll m_L^*$). De stroom I_2 wordt dan toegevoegd aan de bestaande stroom I_1, die naar het L-dal loopt. De drempel V_0 is een directe maat voor het energieverschil tussen het Γ-dal en het L-dal.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

kan nauwkeurig berekend worden; als de druk groter is dan P_1, hangt de ligging van dit niveau sterk af van de druk. In feite zien we de sterke weerstandsdaling bij $V = V_0$ ook nog optreden (bij positieve spanning) als de druk iets lager is dan P_1. De resultaten van deze metingen zijn verzameld in *fig. 8*. Aangezien de spanning V_0 lineair varieert met de druk en aangezien de kleine drukafhankelijkheid van $E_F^{(L)}$ gemakkelijk berekend

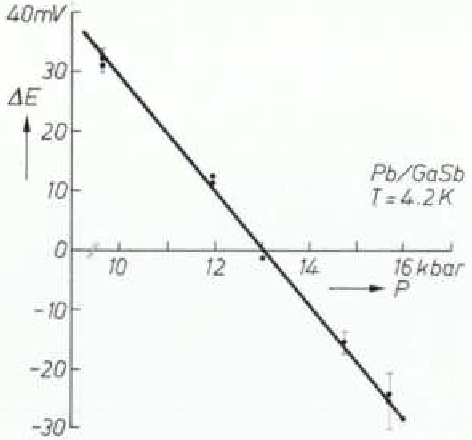


Fig. 8. In deze figuur zijn de resultaten verzameld van alle metingen van de drempelspanning V_0 als functie van de druk. Aangezien $E_F^{(L)}$ vrijwel onafhankelijk is van de druk, kan men uit deze curve ook onmiddellijk het verloop afleiden van de afstand $\Delta E(P)$ tussen het Γ-dal en het L-dal.

kan worden, kunnen we uit de tunnelkarakteristieken $\Delta E(P)$ afleiden. Door extrapolatie kunnen we hieruit de waarde bij $P = 0$ bepalen:

$$\Delta E(P) = \Delta E(0) - \alpha P,$$

met $\alpha = 9{,}6$ meV/kbar en $\Delta E(0) = 100 \pm 5$ meV. De gevonden coëfficiënt α stemt zeer goed overeen met de nieuwste gegevens uit de literatuur. Daarentegen ligt de waarde van $\Delta E(0)$ wel wat hoger dan in de literatuur. In het verleden heeft men hiervoor waarden gevonden van 80 tot 95 meV, afhankelijk van de auteurs en de gebruikte methodes. Hierbij moet worden aangetekend dat, uitgezonderd de tunnelmethode, alle gebruikte experimentele technieken zeer indirect werken en afhangen van de exacte kennis van talrijke parameters. Het tunneleffect vormt thans de meest directe en nauwkeurige methode voor het meten van grootheden als $\Delta E(0)$. Terwijl de drempelwaarde onafhankelijk is van m^* en van E_F, is het duidelijk dat de weerstandswaarde voor $V < V_0$ wèl bepaald wordt door deze grootheden. Men zou derhalve kunnen overwegen deze techniek te gebruiken voor het meten van de effectieve massa.

Tunneleffect met fononen

Het tweede opmerkelijke verschijnsel dat optreedt bij een indirecte kleinste bandafstand (fig. 6, bovenste twee curven), bestaat uit een tweetal sprongen in de differentiële weerstand bij positieve voorspanning. Ze treden op bij spanningen die niet erg afhangen van de drukverhoging; het beste kunnen ze bestudeerd worden door de stroom tweemaal naar de spanning te differentiëren. *Fig. 9a* toont het verloop van deze tweede afgeleide $\mathrm{d}^2I/\mathrm{d}V^2$ voor twee drukken. De onderste curve heeft betrekking op atmosferische druk, waarbij de kleinste bandafstand direct is, en de bovenste geldt voor een van de hoogste drukken die we beschikbaar hadden. Pieken resp. dalen wijzen op stapvormige stijgingen resp. dalingen van het geleidingsvermogen. De onderste curve vertoont twee duidelijke dalen *TA* en *LA* (lood) die kenmerkend zijn voor een supergeleidende loodelektrode [9]. In de bovenste curve zien we bovendien nog vier pieken waarvan de hoogste sterk afhankelijk is van de druk. Deze zijn aangeduid met *TA*, *LA*, *TO* en *LO*. Om de werkelijke hoogte en vorm van de *TA*-piek te kunnen achterhalen, moet de achtergrondcurve worden afgetrokken; deze laatste is in zekere mate gestoord door de aanwezigheid van de „supergeleidende" anomalieën.

Voordat we ingaan op de fysische interpretatie van deze anomale structuren, zullen we eerst de drukafhankelijkheid behandelen van de hoogte van de twee hoofdpieken (*LA* en *TO*). In fig. 9*b* is de relatieve weerstandsdaling $\Delta R/R$ uitgezet die corespondeert met deze hoofdpieken. We zullen op deze figuur nog terugkomen en dan proberen het waargenomen gedrag te verklaren. Intussen is het belangrijk op te merken dat ook dit effect blijkbaar nauw met de bandconfiguratie samenhangt; de sprongen treden namelijk op bij een druk die dicht in de buurt ligt van de kritische waarde P_1 waarbij het Γ-dal volledig onbezet raakt. We kunnen deze verschijnselen verklaren als we aannemen dat de elektronen in een vaste stof niet volkomen vrij zijn, maar een wisselwerking vertonen met hun omgeving, in het bijzonder met de roostertrillingen. Dit laatste mogen we concluderen uit het experimentele feit dat de pieken (in fig. 9*a*) liggen bij spanningen die goed overeenkomen met karakteristieke frequenties van bepaalde fononen.

Laten we nu de dispersierelaties beschouwen, d.w.z. de energie $\hbar\omega_q$ van de fononen als functie van hun golfvector $\boldsymbol{q}$, voor bepaalde kristalrichtingen met sterke symmetrie (*fig. 10*). Deze gegevens verkrijgt men gewoonlijk door niet-elastische verstrooiingsproeven met neutronen. De verschillende krommen in fig. 10 hebben betrekking op de soorten trillingen die de atomen in een eenheidscel van het kristal ten opzichte van elkaar kunnen uitvoeren. Longitudinale (*LO*, *LA*) en transversale (*TO*, *TA*) fononen hebben te maken met samendrukbaarheid, respectievelijk afschuiving. Fononen worden akoestisch (*LA* en *TA*) of optisch (*LO* en *TO*) genoemd

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

al naar de atomen in de eenheidscel in fase of in tegenfase trillen. De maximale amplitude q_0 van de golfvector in een bepaalde richting is afhankelijk van de waarde a van de roosterconstante in die richting: $q_0 = \pi/2a$.

Voor stoffen als germanium en galliumantimonide is bewezen dat de gemiddelde waarde k_L voor de golfvectoramplitude van de elektronen in de L-toestanden, gelijk is aan de maximale waarde q_0 in de bijbehorende (111)-kristalrichting (fig. 2):

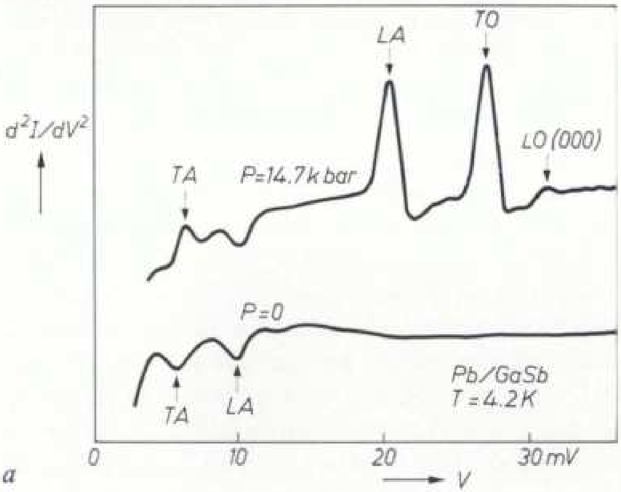
$$k_L(111) = q_0(111).$$

De trilling $q_0(111)$ noemt men een „zonerandfonon" („Brillouin zone-edge phonon"). De spanningen waarbij de pieken in de d^2I/dV^2-curve optreden, blijken goed overeen te stemmen met de energieën van de zonerandfononen van de (111)-richting. Dit wijst erop dat het geleidingsmechanisme waaruit deze scherpe pieken ontstaan, te maken heeft met dalwisselwerkingen, waarbij tunnelende elektronen van de L-dalen naar de Γ-dalen overspringen. De daarbij optredende verandering van impuls leidt tot de emissie van een zonerandfonon:

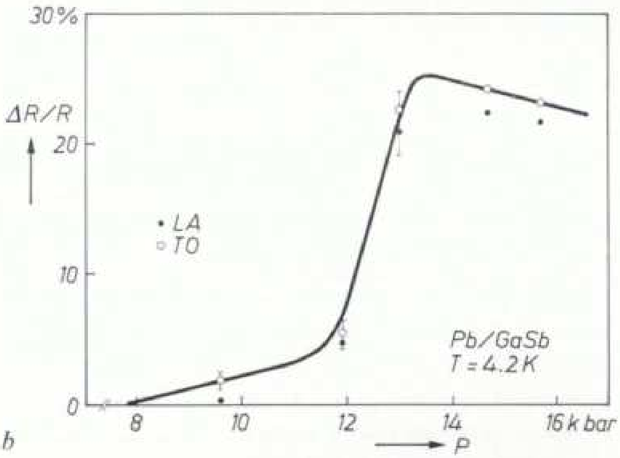
$$k_L(111) - k_\Gamma(0) = q_0(111).$$

Nu we een zeker inzicht hebben verkregen in het verschijnsel dat tot de anomalieën bij positieve voorspanning leidt, willen we de invloed van deze verschijnselen op de tunnelstroom wat nader bezien.

Fig. 11 geeft voor de beide materialen een schematische voorstelling in de $\boldsymbol{k}$-ruimte van het oppervlak van constante energie, links voor het metaal en rechts voor de halfgeleider. Eenvoudigheidshalve hebben we een tweedimensionale $\boldsymbol{k}$-ruimte genomen, waarbij k_x de richting van de tunnelstroom heeft en $k_{||}$ de component van de golfvector voorstelt evenwijdig met het grensvlak. Het oppervlak van constante energie geeft voor



a



b

Fig. 9. *a)* Het verloop van de tweede afgeleide d^2I/dV^2 van de tunnelstroom door een N-GaSb/lood-contact bij een temperatuur van 4,2 K, voor twee verschillende drukken uitgezet tegen de (pos.) aangelegde spanning. De onderste curve heeft betrekking op atmosferische druk; de beide dalen in het verloop van deze curve hebben te maken met de eigenschappen die lood in supergeleidende toestand vertoont. De bovenste curve geldt voor P = 14,7 kbar ($> P_1$); deze vertoont nog vier pieken, die corresponderen met karakteristieke energiewaarden van fononen in GaSb. De beide hoogste pieken (LA en TO) hangen samen met de twee „knikken" in de weerstandscurve bij positieve voorspanning (zie fig. 6). *b)* De relatieve verandering van de differentiële weerstand die optreedt bij deze knikken, uitgezet als functie van de druk. Deze grootheid $\Delta R/R$ neemt sterk toe als P ongeveer gelijk is aan P_1.

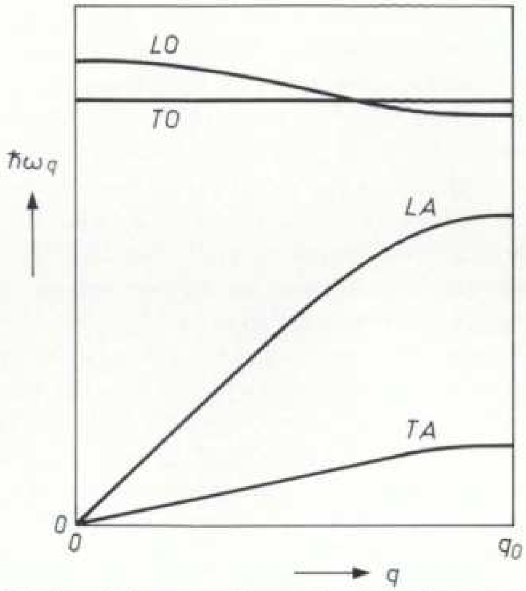


Fig. 10. Schematische voorstelling van het verband tussen energie en impuls in een bepaalde richting ($\hbar\omega_q = f(q)$) voor roostertrillingen in een kristal dat twee atomen per eenheidscel bevat. De vier takken TA, LA, TO, LO beantwoorden aan verschillende soorten trillingen van de roosteratomen ten opzichte van elkaar. q_0 is de maximale waarde van de impuls, die is bepaald door de rand van de eerste Brillouin-zone. De bijbehorende roostertrillingen, met een energie $\hbar\omega_{q_0}$, worden „zonerandfononen" (zone-edge phonons) genoemd.

[9] Deze dalen hangen samen met welbekende anomalieën in de toestandsdichtheid bij supergeleiding en zijn betrekkelijk onafhankelijk van de druk. Ze zouden eigenlijk bij iedere waarde van de druk op overeenkomstige wijze moeten terugkomen. De curve van de tweede afgeleide levert bij atmosferische druk, afgezien van deze beide dalen, geen bijzondere kenmerken op.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

de beide materialen de grens aan van de gebieden in de **k**-ruimte waar toestanden beschikbaar zijn voor elektronen met energiewaarden die bij het tunneleffect optreden. In de halfgeleider is voor $P > P_1$ de kleinste bandafstand indirect; alleen in de L-dalen zijn dan toestanden met de juiste energiewaarden beschikbaar. We hebben de tekening nog verder vereenvoudigd door slechts één L-dal in aanmerking te nemen. Het constante-energieoppervlak is bij dezelfde energiewaarde in het metaal veel groter dan in de halfgeleider. Precies als in de gewone optica bij een spiegel, betekent rechtstreekse transmissie het behoud van $k_{||}$, de golfvectorcomponent evenwijdig met het grensvlak. De tunnelstroom I_1 die bij $P > P_1$ uit de L-dalen komt, is voorgesteld door de pijl *1*. De ermee corresponderende differentiële weerstand van het contact kan berekend worden en de resultaten daarvan zijn in fig. 6 weergegeven door de streeplijnen bij de bovenste twee curven. De transmissiekans die bij pijl *1* in fig. 11 hoort, is alleen afhankelijk van de eigenschappen van de L-dalen;

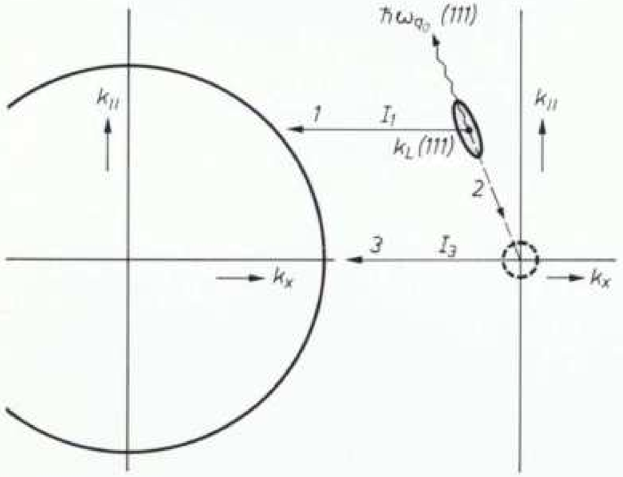


Fig. 11. Een schematisch overzicht van het tweetraps-tunneleffect dat bij halfgeleiders met een indirecte kleinste bandafstand kan optreden door emissie van een zonerandfonon. Voor beide stoffen is het oppervlak van constante energie voor elektrontoestanden in de **k**-ruimte afgebeeld, links voor het metaal, rechts voor de halfgeleider. De **k**-ruimte is voor het gemak tot twee dimensies beperkt. De golfvectorcomponent k_x heeft de richting van de tunnelstroom en $k_{||}$ is de component evenwijdig met het grensvlak, die is verkregen door samenstellen van beide componenten k_y en k_z. Het metaal is isotroop en de golfvector is bij de energieën waarbij het tunneleffect optreedt, zeer groot. In de halfgeleider met indirecte kleinste bandafstand zijn bij deze energiewaarden de beschikbare elektrontoestanden alle geconcentreerd in de L-dalen, die de vorm van een ellipsoïde hebben. Eenvoudigheidshalve is slechts één ellipsoïde afgebeeld. De kleine streepcirkel met $k = 0$ als middelpunt geeft aan dat, wanneer in de halfgeleider bij de tunnelenergie geen stationaire elektrontoestand bestaat, het Γ-dal (met de kleine effectieve massa) bij een iets hogere energie ligt.
Wanneer nu een positieve spanning wordt aangelegd, ontstaat door rechtstreeks tunnelen vanuit de L-dalen een stroom I_1 (pijl *1*). Daarnaast is een tweetrapsproces mogelijk. Bij de eerste stap wekken de elektronen in de L-toestanden een zonerandfonon met energie $\hbar\omega_{q_0}(111)$ op en komen daardoor terecht in het gebied met de kleine effectieve massa rond $k = 0$ (pijl *2*). De tweede stap is dan weer het normale tunnelen naar het metaal (pijl *3*).

doordat de effectieve massa m_L^* groot is, is de kans betrekkelijk klein.

Er bestaat een gebied in de **k**-ruimte waarbij grotere transmissiekansen horen. Dit is het gebied nabij het centrum ($k \approx 0$), dat begrensd wordt door de onderbroken cirkel. Weliswaar is er bij deze energie geen stationaire toestand met golfvector $k \approx 0$ beschikbaar, maar in het Γ-dal, met de kleine effectieve massa, liggen toestanden met een slechts weinig hogere energie. *Bij negatieve spanning* worden dan ook elektronen in de halfgeleider geïnjecteerd waarvan de energie toeneemt met de absolute waarde van de spanning (zie fig. 7). We zien dat, zodra een drempel V_0 overschreden is, de geleiding door rechtstreeks tunnelen naar het Γ-dal domineert. We beschouwen hier echter de situatie bij een *positieve spanning*, en dan is rechtstreeks tunnelen vanuit dit (lege) dal verboden. Het Γ-dal, met zijn kleine effectieve massa, biedt echter toch een grote transmissiekans (pijl *3* in fig. 11), en wel voor elektronen die door een of ander mechanisme eerst uit de bezette L-toestanden naar de Γ-toestanden zijn overgebracht. Dit mechanisme moet de elektronen dan een impuls $k_L(111)$ geven [10]. Het experimentele verschijnsel dat hier ter sprake is, hangt samen met een dergelijk tweetrapsproces waarin de impuls door een zonerandfonon geleverd wordt. Bij zeer lage temperaturen zijn de evenwichtstrillingen van het rooster echter uiterst ineffectief; het is dan zeer onwaarschijnlijk dat een fonon geabsorbeerd wordt. Aangezien de fononen de energie $\hbar\omega_{q_0}$ hebben, zullen de elektronen alleen als de aangelegde spanning V groter is dan $\hbar\omega_{q_0}/e$, genoeg extra-energie bezitten om een fonon te kunnen uitzenden. Voor de stroom met „fonon-assistentie" bestaat dus een drempel: $|eV_t| = \hbar\omega_{q_0}$, wat leidt tot plotselinge weerstandsdalingen bij de spanningen die corresponderen met de energieën van de verschillende groepen fononen. Ofschoon de kans dat elektronen zonerandfononen opwekken klein is, geeft de totale geleiding, afkomstig van het tweetrapsproces, toch een meetbaar effect; de transmissie via pijl *3* in fig. 11 is namelijk veel gemakkelijker dan die via pijl *1*.

Met behulp van dit tweetraps-tunnelproces hebben we voor de eerste keer heel nauwkeurig de energieën van de desbetreffende fononen in GaSb kunnen meten. Nader onderzoek heeft uitgewezen dat echter slechts drie van de pieken in fig. 9*a* aan zonerandfononen kunnen worden toegeschreven (*TA*, *LA*, TO), en dit wekt sterk de indruk dat de beide optische groepen fononen aan de rand van de zone nagenoeg ontaard zijn. De vierde piek, aangegeven met *LO*(000) in fig. 9*a*, hoort bij de longitudinale optische fononen met $q \approx 0$ en wordt hier verder niet besproken.

We zijn nu in staat de drukafhankelijkheid van de amplitude der fononen kwalitatief te verklaren (fig. 9*b*).

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Bij drukken kleiner dan P_1 bevat het Γ-dal elektronen, met andere woorden: er zijn voor alle waarden van de aangelegde spanning Γ-toestanden met de tunnelenergie beschikbaar. Dientengevolge is rechtstreeks tunnelen naar en vanuit het Γ-dal toegestaan. Dit eentrapsproces heeft natuurlijk een grotere waarschijnlijkheid dan het tweetrapsproces, omdat er in het eerste geval geen fonon door het Γ-elektron hoeft te worden opgewekt. De relatieve intensiteit van dit tweetrapsproces gaat snel naar nul als de druk beneden de waarde P_1 daalt. De langzame daling van de amplitude boven P_1 ontstaat doordat het Γ-dal geleidelijk verschuift naar hogere energiewaarden die boven de L-toestanden liggen. We hebben dit kunnen bevestigen met behulp van een aantal experimenten aan N-germanium bij drukverhoging; dit materiaal heeft bij atmosferische druk een indirecte kleinste bandafstand.

De bestudering van de invloed die drukverandering heeft op de plaats en de sterkte van de met fononen gekoppelde anomalieën in de geleiding, verschaft derhalve interessante gegevens over de elektronbandstructuur, over de eigenschappen van de atoomroostertrillingen en ook over de wisselwerking tussen elektronen en fononen.

Ofschoon het tunneleffect ook bij atmosferische druk al belangrijke informatie levert over de materiaaleigenschappen en over het halfgeleider-metaal-contact zelf, zal het duidelijk zijn dat de tunnelmethode door het variëren van een kritische parameter als de druk toch nog zeer veel aan waarde kan winnen. In dit artikel hebben we laten zien hoe men met deze methode zowel de bandstructuur van halfgeleiders kan onderzoeken alsook hun roostertrillingen en de supergeleiding in de metaalelektrode. Ons onderzoek heeft een vrij rechtstreekse bepaling opgeleverd van het energieverschil tussen de diverse potentiaalkuilen van de geleidingsband, en het heeft geleid tot een eerste meting van de energie van zonerandfononen in GaSb. Men mag van deze techniek veel resultaten verwachten bij onderzoek aan nieuwe materialen zoals magnetische halfgeleiders en halfgeleiders met kleine bandafstand, en bij de ternaire verbindingen die gebruikt worden in de opto-elektronica en die, al naar hun samenstelling, een directe of een indirecte kleinste bandafstand bezitten. Halfgeleider-metaal-contacten blijken voor dit soort onderzoek bijzonder geschikt te zijn omdat ze door splijting op zeer betrouwbare en zeer goed reproduceerbare wijze gemaakt kunnen worden; de eigenschappen van deze contacten kunnen hierdoor kwantitatief verklaard worden [11].

[10] We merken op dat de afwezigheid van beschikbare toestanden bij $k \approx 0$ en bij de tunnelenergie dit tweetrapsproces niet verbiedt. Dit komt doordat de tussenstap een korte levensduur heeft en dus een slecht gedefinieerde energie bezit.

[11] P. Guétin en G. Schréder, J. appl. Phys. **42**, 5689, 1971 (no. 13).

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

 Philips techn. T. 32, 376-386, 1971, no. 10/11/12

Het laterale skineffect in platte geleiders

V. Belevitch

Inleiding

Het klassieke skineffect is een welbekend verschijnsel, onlangs nog uitvoerig in dit tijdschrift besproken door H. B. G. Casimir en J. Ubbink [1]. Daarbij wordt in een draad met cirkelvormige doorsnede de radiale verdeling van de stroomdichtheid gegeven door een Besselfunctie waarvan het argument evenredig is met de wortel uit de frequentie. Bij gelijkstroom is de stroom homogeen verdeeld, maar bij hoge frequenties is hij geconcentreerd in een laag aan de buitenkant ter dikte

$$\delta = \sqrt{(2/\mu\varrho\omega)}, \qquad (1)$$

de zg. *skin-dikte* (μ is de permeabiliteit, ϱ het geleidingsvermogen, ω de hoekfrequentie; praktische elektromagnetische eenheden). Dienovereenkomstig stijgt bij toenemende frequentie de weerstand R per lengte-eenheid aanvankelijk evenredig met het kwadraat van de frequentie, maar uiteindelijk omgekeerd evenredig met de skin-dikte, dus evenredig met de wortel uit de frequentie.

Kwalitatief geldt hetzelfde voor geleiders van willekeurige doorsnede en met een gladverlopende rand; daarbij is de neiging van de stroom om zich aan het oppervlak te concentreren het sterkst daar waar het oppervlak het sterkst is gekromd. Zo zal in een geleider met elliptische doorsnede de stroomdichtheid bij hoge frequentie aan de uiteinden van de lange as groter zijn dan aan de uiteinden van de korte as.

Dit effect is sterk als de excentriciteit van de ellips groot is; een interessant probleem is het limietgeval van een zeer platte elliptische cilinder (*fig. 1a*). Van groot praktisch belang is natuurlijk het geval van de platte rechthoekige band (fig. 1*b*), die veel wordt gebruikt bij gedrukte bedradingen en microgolftoepassingen. Nu is echter alleen het elliptische geval analytisch te behandelen, het rechthoekige geval moet langs numerieke weg worden aangepakt. Daarom is het goed beide gevallen te beschouwen, in de hoop uit de theorie van de platte ellips enige resultaten af te leiden die, althans kwalitatief, ook voor de platte rechthoek gelden.

Het probleem van het skineffect in platte geleiders wordt hierdoor gekenmerkt dat er twee afmetingen in voorkomen, een dikte $2b$ en een breedte $2a$, van verschillende orde van grootte:

$$b \ll a. \qquad (2)$$

Daardoor splitst het probleem zich automatisch in twee bijna los van elkaar staande problemen. Bij frequenties die zo laag zijn dat

$$\delta \gg b \qquad (3)$$

is, is de stroom nog homogeen over de dikte verdeeld, zodat het enige probleem is hoe de totale stroom per cm breedte, de zg. lineaire stroomdichtheid $i(x)$ afhangt van de breedtecoördinaat x. Dit is het probleem van het *laterale skineffect*, het onderwerp van dit artikel. Bij de veel hogere frequenties waarvoor $\delta \ll b$ is, treedt over de gehele omtrek het normale, bekende skineffect op.

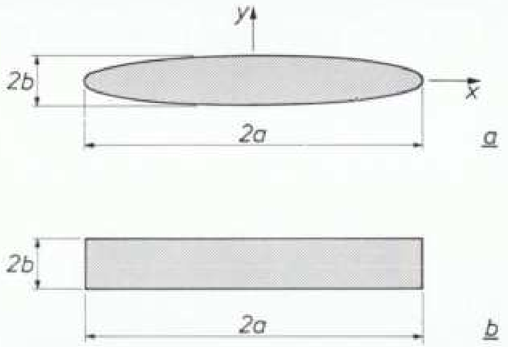


Fig. 1. *a*) Cilindrische geleider met elliptische doorsnede. *b*) Cilindrische geleider met rechthoekige doorsnede.

De opzet van het artikel is als volgt. Het eerste hoofdstuk geeft een kwalitatieve beschrijving van het laterale skineffect in platte geleiders en de daarmee gepaard gaande toeneming van de weerstand bij stijgende frequentie. De volledige wiskundige analyse wordt tot het laatste hoofdstuk bewaard; enkele wiskundige opmerkingen vooraf zijn echter onontbeerlijk en worden in het tweede hoofdstuk gemaakt. Dan volgt een samenvatting van de volledige analytische resultaten voor de platte ellips — alle oorspronkelijk, voor zover ons bekend — en tevens van de betrekkelijk schrale reeds bestaande informatie over de platte rechthoek, alsmede een nieuw daarop betrokken detailresultaat. In het

Prof. dr. V. Belevitch, directeur van MBLE Laboratoire de Recherches, Brussel, België.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

daarop volgende hoofdstuk wordt, zowel voor de ellips als voor de rechthoek, de impedantie gekarakteriseerd door middel van zijn polen en nulpunten, hetgeen dieper inzicht in het gedrag van de impedantie verschaft. Ten slotte wordt in het Aanhangsel een poging ondernomen het hoogfrequentgedrag van de impedantie voor de rechthoek te analyseren.

Het laterale skineffect

Zoals in de inleiding is opgemerkt, kan het skineffect in platte geleiders worden gesplitst in een lateraal probleem en het normale skineffect. Het laterale probleem is voor geleiders zonder scherpe kanten (zoals een holle elliptische cilinder) nog weer anders van aard dan voor geleiders met scherpe kanten. Aangezien voorwaarde (2) zowel voor de ellips als voor de rechthoek scherpe kanten (punten met oneindig sterke kromming) bij $x = \pm a$ impliceert, is in beide gevallen de stroomconcentratie bij de uiteinden het belangrijkst; deze leidt voor ellips en rechthoek tot verschillende wetmatigheden voor de weerstandstoeneming. In dit hoofdstuk beschouwen we achtereenvolgens (a) platte geleiders in het algemeen, (b) platte geleiders zonder scherpe kanten, (c) platte geleiders met scherpe kanten.

In plaats van de drie afmetingen a, b en δ introduceren we gemakshalve de twee dimensieloze verhoudingen

$$\frac{\sqrt{(ab)}}{\delta} \text{ en } \frac{b}{a}. \qquad (4)$$

Het laterale probleem is gedefinieerd door de voorwaarde $b/a = 0$, grensgeval van voorwaarde (2). Het asymptotische gedrag van het laterale skineffect zelf doet zich voor bij zodanige frequenties dat $\sqrt{(ab)}/\delta \gg 1$ is. De lineaire stroomdichtheid $i(x)$ en de uitwendige elektromagnetische velden hebben in dit grensgeval hf.-waarden die onafhankelijk zijn van de frequentie. Dit gedrag noemen we *lateraal-asymptotisch*; samengevat doet het zich bij zodanige frequenties voor dat

$$b \ll \delta \ll \sqrt{(ab)} \qquad (5)$$

is, voorwaarden die niet met (2) in strijd zijn.

Als de frequentie zo ver toeneemt dat (3) niet langer geldt, begint de echte stroomdichtheid (A/cm^2) te variëren langs de diktecoördinaat, maar de lineaire dichtheid (A/cm) en de externe velden behouden hun lateraal-asymptotische waarden, gezien de genoemde splitsing van het probleem in lateraal en normaal skineffect.

Platte geleiders vertonen een geheel ander variatiepatroon van de weerstand met de frequentie dan normale „dikke" geleiders. Er zijn namelijk twee verschillende fasen van toeneming (het laterale en het normale effect), waarvoor verschillende wetmatigheden gelden en die worden gescheiden door een groot, door (5) gegeven frequentie-interval, waarin het laterale effect reeds zijn asymptotisch gedrag vertoont terwijl het normale effect nog geen rol speelt. De lineaire stroomdichtheid en de externe velden bereiken hun asymptotische waarden in de eerste fase en blijven onveranderd in de tweede fase.

Voor een platte geleider zonder scherpe kanten, zoals een platte, holle elliptische cilinder met matige excentriciteit, is de lateraal-asymptotische waarde van de lineaire stroomdichtheid eindig in ieder punt, en ook de lateraal-asymptotische weerstand is eindig. De variatie van de weerstand met de frequentie verloopt derhalve kwalitatief zoals geschetst is in *fig. 2*: eerst een laterale fase AB, gevolgd door een lang, door (5) bepaald, interval BC waarin de weerstand zijn lateraal-asymptotische waarde heeft; daarna de fase CD van

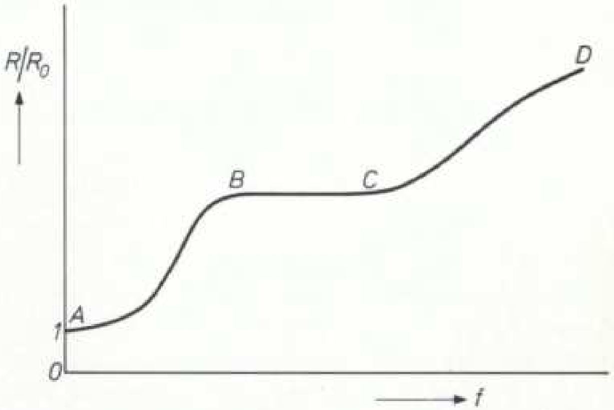


Fig. 2. De weerstand R als functie van de frequentie f voor een platte, holle elliptische cilinder: AB is de laterale fase, BC de lateraal-asymptotische waarde, CD de fase van het normale skineffect. R_0 is de gelijkstroomwaarde van R. De kromme geldt slechts kwalitatief.

het normale skineffect, waarin de weerstand ten slotte evenredig met de wortel uit de frequentie toeneemt.

Hoewel het probleem van de platte holle elliptische cilinder een tamelijk academisch karakter heeft, kan een korte bespreking van de lateraal-asymptotische fase BC ervan verhelderend werken. Aangezien het externe magneetveld voldoet aan de vergelijking van Laplace en geen component loodrecht op de ellips heeft, zijn de krachtlijnen homofocale ellipsen. De lineaire stroomdichtheid (langs de rand van de ellips) wordt direct gegeven door de discontinuïteit van de tangentiële component van het magneetveld, en is dus omgekeerd evenredig met de afstand tussen twee naburige ellipsen. In het bijzonder is de lineaire stroomdichtheid onafhankelijk van de dikte $2h$ van de cilinder, ook als deze van plaats tot plaats verschilt. Daaren-

[1] H. B. G. Casimir en J. Ubbink, Het skineffect, Philips techn. T. 28, 173-185, 206-211, 376-392, 1967.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

tegen hangt de lateraal-asymptotische weerstand (de ordinaat van *BC* in fig. 2) af van de dikte, omdat de echte stroomdichtheid $i/2h$ (A/cm²) in een oppervlakte-element $2h\mathrm{d}s$ ($\mathrm{d}s$ is het lijnelement langs de rand) een vermogen $(i/2h)^2\, 2h\mathrm{d}s$ produceert, zodat het totaal gedissipeerde vermogen

$$\int (i/2h)^2\, 2h\mathrm{d}s \qquad (6)$$

is, dus in het algemeen afhankelijk van h. Indien in het bijzonder de dikte evenredig met $i(x)$ wordt gekozen, is $i/2h$ in (6) een constante, evenals bij gelijkstroom, zodat R de asymptotische waarde R_0 heeft. Aangezien de weerstand bij stijgende frequentie niet kan afnemen, moet hij, bij dit verloop van de dikte, in de gehele laterale fase constant blijven. Een holle elliptische cilinder heeft juist dit dikteverloop indien de geleider zich tussen twee gelijkvormige ellipsen bevindt. In een dergelijke geleider wordt de toeneming van de lineaire stroomdichtheid naar de uiteinden van de lange as toe precies gecompenseerd door de toegenomen dikte, zodat de echte stroomdichtheid gelijk blijft. Een triviaal voorbeeld is een holle cirkelvormige cilinder van constante dikte; daarin treedt geen lateraal skineffect op.

Deze beschouwing, die tot het weerstandsgedrag van fig. 2 leidt, is alleen geldig voor een platte geleider zonder scherpe kanten, en niet voor een platte band, hetzij rechthoekig hetzij elliptisch. De lateraal-asymptotische lineaire stroomdichtheid wordt daarbij namelijk oneindig aan de kanten ($x = \pm a$), zodat de dissipatie (6), en dus ook de weerstand, oneindig is voor ieder verloop van de dikte $h(x)$, tenzij h ook oneindig zou worden aan de kanten, maar dat is strijdig met de veronderstelling dat de geleider plat is. De lateraal-asymptotische lineaire stroomdichtheid voor een band van willekeurige dikte is

$$i(x) = \frac{2i_0}{\pi\sqrt{(1 - x^2/a^2)}}, \qquad (7)$$

waarin i_0 de gemiddelde lineaire stroomdichtheid is:

$$i_0 = I/2a; \qquad (8)$$

I is de totale stroom. Uitdrukking (7) is reeds lang bekend [2], maar zal in dit artikel opnieuw worden afgeleid. Voor een rechthoekige doorsnede varieert de lineaire stroomdichtheid van een van x onafhankelijke dichtheid in het geval van gelijkstroom, gegeven door (8), tot de lateraal-asymptotische dichtheid (7) bij hoge frequenties. Voor de elliptische doorsnede is echter de *ware* stroomdichtheid $i/2h$ in het geval van gelijkstroom overal gelijk. Aangezien de geleider van fig. 1*a* een variabele dikte $2h(x)$ heeft, waarbij

$$h = b\sqrt{(1 - x^2/a^2)} \qquad (9)$$

is, is de lineaire dichtheid bij gelijkstroom niet onafhankelijk van x, maar wordt gegeven door

$$\frac{i(x)}{i_0} = \frac{4}{\pi}\sqrt{(1 - x^2/a^2)}; \qquad (10)$$

in het elliptische geval varieert de lineaire dichtheid derhalve veel sterker als de frequentie vanaf nul onbegrensd toeneemt.

Aangezien de lateraal-asymptotische weerstand van een platte band oneindig is, moet de weerstand kwalitatief met de frequentie variëren op de wijze die is geschetst in *fig. 3*. Aan het einde van de laterale fase *AB* bereikt de weerstand zijn lateraal-asymptotisch gedrag *BCE*, waarvan hij weer gaat afwijken zodra het normale skineffect een rol gaat spelen (*CD*). De kromme *BCE* gaat op een vooralsnog onbekende wijze naar oneindig, maar in elk geval langzamer dan volgens de wortel uit de frequentie, aangezien hij in de fase *CD* door het normale skineffect wordt overheerst. Ten slotte zij opgemerkt dat zowel het laterale gedrag *AB* als het lateraal-asymptotische gedrag *BCE* voor rechthoek en ellips verschillend zijn, maar het gedrag aan het einde van fase *CD* (de evenredigheid met de wortel uit de frequentie) is voor beide gelijk.

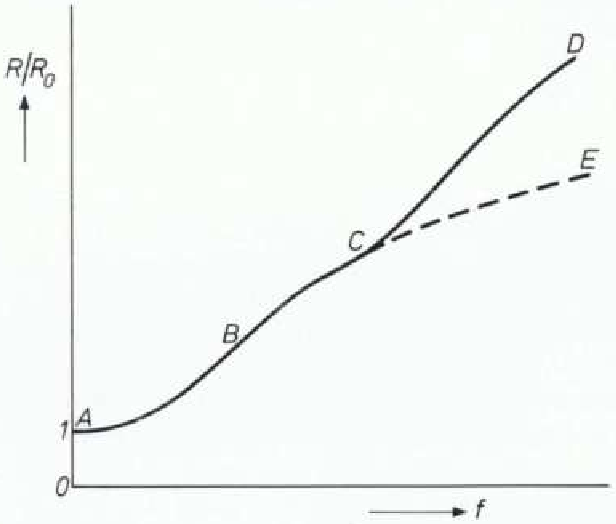


Fig. 3. Als fig. 2, maar nu voor een platte band. De kromme *BCE* geeft het lateraal-asymptotische gedrag weer.

Aspecten van de mathematische behandeling

Ten einde dubbel werk bij de analyse van de elliptische en de rechthoekige band te vermijden, stellen we in beide gevallen de dikte van de band ter plaatse x voor door $2h(x)$; bij de rechthoek is h gelijk aan b, bij de ellips wordt h gegeven door (9). Als een voor beide gevallen bruikbare genormeerde frequentie introduceren we de variabele

$$k = \frac{\mathrm{j}\omega\mu}{4\pi R_0} = \mathrm{j}\Omega. \qquad (11)$$

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Deze is evenredig met het kwadraat van de eerste van de twee parameters (4). Voor ellips en rechthoek vinden we:

		ellips	rechthoek
gelijkstroomweerstand per lengte-eenheid	R_0	$\dfrac{1}{\pi\varrho ab}$	$\dfrac{1}{4\varrho ab}$
genormeerde frequentie	k	$\dfrac{j\omega\mu\varrho ab}{4}$	$\dfrac{j\omega\mu\varrho ab}{\pi}$

(12)

De vergelijking voor het klassieke skineffect in een dikke cilinder is

$$\Delta u + \lambda u = 0, \tag{13}$$

met

$$\lambda = -\varrho\mu j\omega. \tag{14}$$

Hierin is u de axiale component van elektrisch veld, stroomdichtheid of vectorpotentiaal en Δ de tweedimensionale Laplace-operator. In het geval van de cirkelvormige cilinder zijn de magnetische krachtlijnen bij alle frequenties concentrische cirkels, zodat er geen flux is van de geleider naar het omgevende diëlektricum of omgekeerd. Dientengevolge kan de interne stroomverdeling op zichzelf worden bestudeerd als oplossing van (13) met cirkelsymmetrie, en het externe veld wordt gewoonlijk verwaarloosd. Daarentegen heeft het magneetveld voor elliptische en rechthoekige doorsneden een normale component die niet nul is en die de geleider binnendringt (behalve bij hoge frequenties), zodat het interne en het externe probleem niet te scheiden zijn. Voor het diëlektricum ($\varrho = 0$ dus $\lambda = 0$) gaat vergelijking (13) over in de vergelijking van Laplace. De externe en de interne oplossingen zijn aan elkaar gekoppeld door middel van randvoorwaarden. Voor platte geleiders blijven alleen de vergelijking van Laplace en de randvoorwaarden over, hetgeen een aanzienlijke vereenvoudiging betekent.

Een stroom I door een cilindrische geleider produceert op grote afstand D van het middelpunt een magneetveld met tangentiële component $I/2\pi D$. De in een cilinder met straal D geproduceerde flux per lengte-eenheid is dus evenredig met ln D, en derhalve onbegrensd als D naar oneindig gaat. Wegens deze moeilijkheid kan de retourleiding niet worden weggedacht. Nu hangt het externe probleem af van de plaats van de retourleiding. In het algemeen veronderstelt men dat de retourleiding een concentrische cilinder is met een grote, doch eindige straal D. Daardoor wordt de invloed van de retourleiding op de stroomverdeling in de centrale geleider verwaarloosbaar, maar de willekeurige constante D verschijnt in alle uitdrukkingen die de inductie bevatten.

Een geleider waarvan de impedantie $Z = R + j\omega L$ door het skineffect wordt bepaald, komt overeen met een RL-keten met verdeelde elementen. In het bijzonder leert de netwerktheorie dat de inductie L een monotoon afnemende functie van de frequentie is die zijn minimum L_∞ bereikt bij frequentie oneindig. Aangezien er bij frequentie oneindig geen inwendig magneetveld is, is L_∞ de enige natuurlijke definitie van de uitwendige inductie. Bovendien is $Z - j\omega L_\infty$ dan een impedantie met minimale reactantie en verdwijnt de constante D in dit verschil. In het vervolg berekenen we steeds de gereduceerde impedantie

$$z = \frac{Z - j\omega L_\infty}{R_0}, \tag{15}$$

waarin R_0 de gelijkstroomweerstand (12) is. De uitwendige inductie L_∞ hangt samen met de capaciteit van het geleiderpaar per lengte-eenheid volgens $CL_\infty = 1/c^2$, waarbij c de lichtsnelheid is. Voor een platte band (ellips of rechthoek) geldt:

$$L_\infty = \frac{\mu}{4\pi} \ln \frac{2D}{a}. \tag{16}$$

Samenvatting van de resultaten

De platte elliptische band

Voor de platte elliptische band is de gereduceerde impedantie (15):

$$z = \frac{J_k(k)}{J_k'(k)}, \tag{17}$$

waarin $J_k(k)$ de Bessel-functie van de eerste soort is met als orde en argument de door (11) gegeven grootheid k. De noemer van (17), $J_k'(k)$, is de afgeleide naar het argument (en niet naar de orde), en is dus $dJ_k(s)/ds$ voor $s = k$.

De lineaire stroomdichtheid kan in de hulpvariabele

$$u = \arccos x/a \tag{18}$$

op drie equivalente manieren worden uitgedrukt:

$$\frac{i}{i_0} = \frac{2}{\pi|\sin u|}\left[1 - \frac{2}{kJ_k'(k)} \Sigma\, nJ_{n+k}(k)\cos 2nu\right], \tag{19}$$

$$\frac{i}{i_0} = \frac{4|\sin u|}{\pi J_k'(k)}\left[J_k(k) + 2\Sigma\, J_{n+k}(k)\cos 2nu\right], \tag{20}$$

$$\frac{i}{i_0} = \frac{4}{\pi J_k'(k)} \Sigma \left[J_{k+n-1}(k) - J_{k+n}(k)\right] \sin(2n-1)u, \tag{21}$$

[12] H. Kaden, Arch. Elektrontechik 28, 818, 1934.

Digitized in Heiloo the Netherlands Not for sale!

waarbij steeds gesommeerd wordt over $n = 1, 2, \ldots \infty$, en waarin i_0 door (8) wordt gegeven. Voor k oneindig blijft in (19) alleen de eerste term over, die met behulp van (18) in (7) overgaat. De reeks (19) is echter divergent in de eindpunten $x = \pm a$, die corresponderen met $u = 0$ of π. Daarentegen zijn (20) en (21) convergent. Vergelijking (20) gaat over in (10) voor $k = 0$.

De uitdrukking (17) kan in een kettingbreuk worden ontwikkeld:

$$z = 1 + \cfrac{1}{\cfrac{2}{k} + \cfrac{1}{1 + \cfrac{1}{\cfrac{4}{k} + \ldots}}} \tag{22}$$

De opeenvolgende noemers hiervan zijn afwisselend 1 en $2n/k$ ($n = 1, 2, \ldots$). Dientengevolge kan de impedantie (17) door het vervangingsschema van *fig. 4* gerepresenteerd worden. Verwaarlozing van de derde- en hogeregraadstermen in (22) levert:

$$z = 1 + \frac{k}{2} - \frac{k^2}{4}. \tag{23}$$

De asymptotische waarde van (17) voor hoge frequenties is [3]:

$$z = C\,k^{1/3}, \tag{24}$$

waarin de factor C wordt gegeven door:

$$C = \frac{\Gamma(1/3)}{2^{1/3}3^{1/3}\,\Gamma(2/3)} = 1{,}088\ldots. \tag{25}$$

Gebruik makend van de definitie van Ω (zie verg. 11) vinden we:

$$z = (0{,}942\ldots + \mathrm{j}.0{,}544\ldots)\Omega^{1/3}. \tag{26}$$

In *fig. 5* zijn het reële en het imaginaire gedeelte van de uitdrukking (17) als functie van Ω uitgezet, te zamen met de raaklijnen in $\Omega = 0$ volgens (23) en de asymptotische waarden volgens (26). De numerieke berekening van (17) was gebaseerd op de ontwikkeling (22), zodanig afgebroken dat 20 dan wel 30 RL-secties in fig. 4 in rekening waren gebracht; dit leverde voor de resultaten binnen het gebied $|k| < 10$ geen verschil van betekenis.

De platte rechthoekige band

Vergeleken met deze volledige analytische resultaten voor de platte ellips is er aangaande de platte rechthoek erg weinig bekend. Via een numeriek onderzoek van de integraalvergelijking voor de lineaire stroomdichtheid hebben Belevitch c.s. [4] benaderingen voor de impedantie bij lage frequenties verkregen. Anderzijds hebben Silvester [5] en Beccari en Ronca [6] via een zuiver numerieke aanpak van het probleem (met verschillende methodes) een weerstandsgedrag gevonden dat de vroegere metingen van Kennelly en Affel [7] bevestigt. Uit al deze resultaten blijkt dat de relatieve weerstandstoeneming, $\mathrm{Re}\,(z - 1)$, voor de ellips ongeveer twee maal zo groot is als voor de rechthoek. Het nieuwe detailresultaat aangaande de platte rechthoek dat wij in dit artikel naar voren willen brengen is de volgende analytische uitdrukking voor de weerstand, een verhouding van twee oneindige determinanten:

$$z = \Delta_{11}/\Delta, \tag{27}$$

waarbij Δ de determinant is van de symmetrische matrix

$$\left[\begin{array}{c|cccc} 1 & -\frac{2}{1\times3} & -\frac{2}{3\times5} & -\frac{2}{5\times7} & \cdots \\ \hline -\frac{2}{1\times3} & 2\left(1+\frac{1}{k}-\frac{1}{3\times5}\right) & -2\left(\frac{1}{1\times3}+\frac{1}{5\times7}\right) & -2\left(\frac{1}{3\times5}+\frac{1}{7\times9}\right) & \cdots \\ -\frac{2}{3\times5} & -2\left(\frac{1}{1\times3}+\frac{1}{5\times7}\right) & 2\left(1+\frac{2}{k}-\frac{1}{7\times9}\right) & -2\left(\frac{1}{1\times3}+\frac{1}{9\times11}\right) & \cdots \\ -\frac{2}{5\times7} & -2\left(\frac{1}{3\times5}+\frac{1}{7\times9}\right) & -2\left(\frac{1}{1\times3}+\frac{1}{9\times11}\right) & 2\left(1+\frac{3}{k}-\frac{1}{11\times13}\right) & \cdots \\ \cdots & & & & \end{array}\right], \tag{28}$$

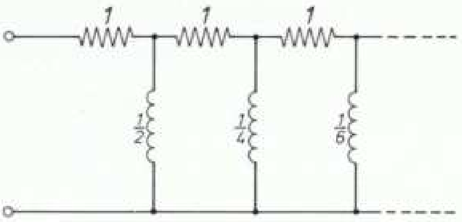


Fig. 4. Vervangingsschema voor de impedantie (17), gebaseerd op de ontwikkeling (22) van (17).

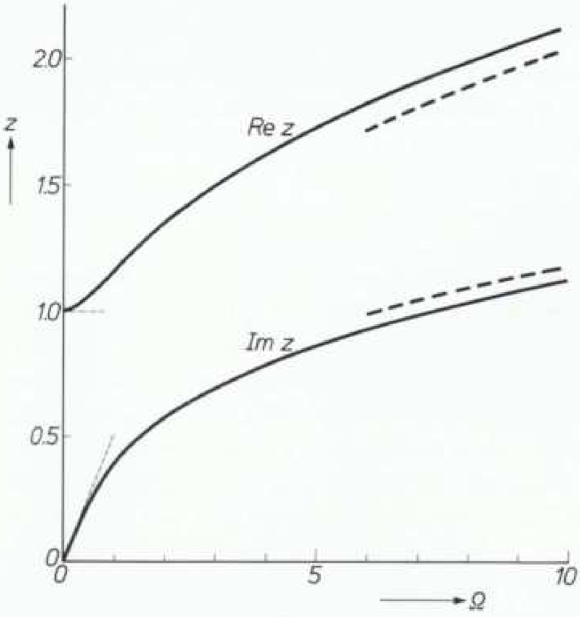


Fig. 5. Het reële en het imaginaire gedeelte van de functie $z = J_k(k)/J_k'(k)$, met $k = \mathrm{j}\Omega$. De streeplijnen zijn de raaklijnen voor $\Omega = 0$ en de asymptoten voor grote Ω.

www.freeservicemanuals.info 11/14/2023
Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

en Δ_{11} de met stippellijnen aangegeven hoofdminor. Matrixelementen op een evenwijdige van de hoofddiagonaal hebben gelijke eerste termen, die op een evenwijdige van de nevendiagonaal gelijke laatste termen. De resultaten (27) en (28) worden in het volgende hoofdstuk en in het Aanhangsel nader besproken.

Polen en nulpunten

Voor de door (15) gedefinieerde impedantie $z(k)$ geldt:

$$z(0) = 1. \tag{29}$$

Voorts heeft $z(k)$ een RL-karakter, zodat zijn nulpunten $k_1, k_2, \ldots$ en zijn polen $k_1', k_2', \ldots$ negatief reëel zijn en elkaar afwisselen:

$$0 < -k_1 < -k_1' < -k_2 < -k_2' < \ldots . \tag{30}$$

In dit hoofdstuk onderzoeken wij de verdeling van polen en nulpunten, zowel voor de ellips als voor de rechthoek, met de bedoeling meer informatie te verkrijgen over het impedantiegedrag, in het bijzonder over het asymptotisch gedrag van de rechthoek.

Fysisch gesproken karakteriseren polen en nulpunten zeker geen trillingen; misschien de wijze van uitdempen (met verschillende randvoorwaarden). Zij corresponderen met vrije oplossingen van de skineffectvergelijking, en dus met eigenwaarden λ van de Laplace-operator in (13). De asymptotische verdeling van deze eigenwaarden is, in verband met statistisch-thermodynamische problemen, in de mathematische fysica reeds herhaaldelijk bestudeerd; in het bijzonder heeft men gevonden [8] dat de n'de eigenwaarde λ_n van (13) voor een tweedimensionaal gebied met oppervlakte S onder vrij algemene randvoorwaarden asymptotisch naar $4\pi n/S$ gaat. Met de vergelijkingen (14) en (11) en de betrekking $R_0 = 1/\varrho S$ volgt daaruit dat de asymptotische waarde van de n'de pool of het n'de nulpunt $-n$ is, hetgeen betekent dat

$$k_n = -n - \alpha_n ; k_n' = -n - \beta_n \quad (n = 1, 2, \ldots \infty), \tag{31}$$

waarbij α_n/n en β_n/n naar nul gaan als n onbegrensd toeneemt. Bovendien legt het alterneren van de polen en de nulpunten (30) aan de afwijkingen α_n en β_n de volgende beperking op:

$$0 < \beta_n - \alpha_n < 1. \tag{32}$$

Hoewel de asymptotische verdeling (31) onafhankelijk is van de vorm van de doorsnede, en dus ook geldt voor een cirkelvormige doorsnede met straal a, worden in dit geval, vanwege de cirkelsymmetrie, niet alle eigentrillingen door de stroom aangeslagen, zodat slechts een kleine deelverzameling van de eigenwaarden (corresponderend met een eendimensionaal probleem, waarbij λ_n asymptotisch nadert tot $n^2\pi^2/a^2$) als polen en nulpunten van de impedantie verschijnt.

Aangezien α_n/n en β_n/n zeker naar nul gaan als α_n en β_n tot constante waarden naderen, is het interessant het geval te bespreken waarbij α_n en β_n constant zijn in (31). De functie die aan (29) voldoet is dan:

$$\frac{\Gamma(\alpha+1)\,\Gamma(k+\beta+1)}{\Gamma(\beta+1)\,\Gamma(k+\alpha+1)}. \tag{33}$$

Deze nadert, volgens de benadering van Stirling van de gammafunctie voor grote $|k|$, asymptotisch tot

$$\frac{\Gamma(\alpha+1)}{\Gamma(\beta+1)} k^{\beta-\alpha}. \tag{34}$$

Als α_n en β_n niet constant zijn, maar voldoende snel tot constante waarden α en β naderen, is de asymptotische uitdrukking van de impedantie nog wel van de vorm $Ck^{\beta-\alpha}$, maar met een coëfficiënt C die verschilt van die van (34); deze laatste, reeds in (33) optredende coëfficiënt volgt uit voorwaarde (29). Het asymptotisch verloop kan echter geheel anders worden als de variatie van α_n of β_n zeer langzaam is, zoals het voorbeeld van de logaritmische afgeleide van de gammafunctie, $\Psi(k) = \Gamma'(k)/\Gamma(k)$, laat zien. Voor de functie $\Psi(k+1)$ geldt $\beta_n = 0$, terwijl α_n naar nul gaat [9] volgens $-1/\ln n$. Hoewel de asymptotische uitdrukking (34) de waarde 1 heeft voor $\beta - \alpha = 0$, gaat de functie Ψ volgens $\ln k$ naar oneindig.

Wat de elliptische impedantie (17) betreft, weet men [10] dat het n'de nulpunt van $J_k(k)$ asymptotisch naar $-n + \frac{1}{6}$ gaat. Aangezien voor de ellips de exponent $\beta - \alpha$ in (34) volgens (24) gelijk is aan $\frac{1}{3}$, is het te verwachten dat de n'de pool asymptotisch naar $-n - \frac{1}{6}$ gaat. Dit wordt bevestigd door *tabel I*, die het resultaat is van een numerieke berekening waarbij 20 dan wel 30 secties in het schema van fig. 4 in rekening zijn gebracht.

Tabel I. Nulpunten (k_n) en polen (k_n') van de functie $J_k(k)/J_k'(k)$.

n	k_n	k_n'
1	−0,83752	−1,11712
2	−1,83490	−2,13294
3	−2,83422	−3,14014
4	−3,83393	−4,14441
5	−4,83377	−5,14728
6	−5,83368	−6,14937

Voor de rechthoek hebben wij de nulpunten en polen volgens (27) en (28) berekend, met matrices die bij orde 30 dan wel 40 afgebroken zijn. De resultaten ziet men in *tabel II*, die het asymptotisch gedrag (31) bevestigt. Dit gedrag vindt men, ondanks andere norme-

[3] M. Abramowitz en I. A. Stegun, Handbook of mathematical functions, Nat. Bur. Stand., Washington 1964, blz. 368.
[4] V. Belevitch, P. Guéret en J. C. Liénard, Rev. HF 5, 109, 1962.
[5] P. Silvester, Proc. IEEE 54, 1147, 1966.
[6] C. Beccari en C. Ronca, Elettrotecnica 56, 607, 1969.
[7] A. E. Kennelly en H. A. Affel, Proc. I.R.E. 4, 523, 1916.
[8] R. Courant en D. Hilbert, Methods of mathematical physics, Vol. I, Interscience, New York 1953, blz. 432.
[9] Zie Abramowitz en Stegun [3], blz. 259.
[10] E. Jahnke, F. Emde en F. Lösch, Tafeln höherer Funktionen, 6e ed., Teubner, Stuttgart 1960, blz. 154.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Tabel II. Nulpunten (k_n) en polen (k_n') van de impedantie van een platte rechthoekige geleider.

n	k_n	k_n'
1	− 0,87914	− 0,9716
2	− 2,06221	− 2,1591
3	− 3,10107	− 3,2156
4	− 4,12081	− 4,2496
5	− 5,13360	− 5,2742
6	− 6,14298	− 6,2935
7	− 7,1504	− 7,3093
8	− 8,1565	− 8,3227
9	− 9,1618	− 9,3341
10	−10,1664	−10,344
11	−11,1705	−11,352
12	−12,1743	−12,358
13	−13,1777	−13,360
14	−14,1809	
15	−15,1838	

ringen dan (11) en (15), ook in de resultaten van Silvester [5] terug; zijn tabel, die de nulpunten (maar niet de polen) geeft van $z + k \ln 2$ als functie van de variabele $4k$, is in overeenstemming met onze resultaten.

Hoewel de eigenfrequenties voor rechthoek en ellips asymptotisch gelijksoortig verdeeld zijn, vertonen de afwijkingen α_n en β_n op kleine schaal markante systematische verschillen, zoals blijkt uit *fig. 6*. In tegenstelling met de snelle benadering van de asymptoten bij de ellips, vertonen bij de rechthoek de afwijkingen een logaritmisch verloop, waarbij het verschil $\beta_n - \alpha_n$ toeneemt. Deze toeneming kan volgens (32) niet onbeperkt doorgaan, zodat wij het asymptotische gedrag van de afwijkingen nog lang niet hebben bereikt. Daar echter de voortzetting van het logaritmische verloop met de hellingen van fig. 6 pas ongeveer bij $n = 10^8$ met (32) in strijd komt, is er weinig hoop het echte asymptotische gedrag via numerieke berekeningen te bepalen.

Het feit dat we de asymptotische uitdrukking (24) voor de impedantie van de ellips hebben kunnen vinden, is te danken aan het bestaan van de gesloten

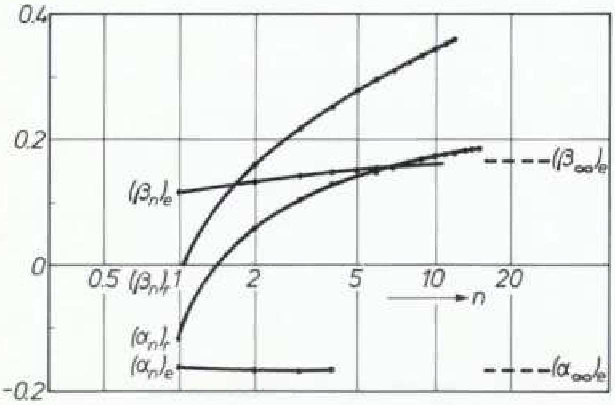


Fig. 6. De afwijkingen van n van de nulpunten en de polen van de genormeerde impedantie, voor een geleider met elliptische doorsnede (indices e) en met rechthoekige doorsnede (indices r): $\alpha_n = k_n - n$ en $\beta_n = k_n' - n$. De streeplijnen zijn de asymptotische waarden $(\alpha_\infty)_e$ en $(\beta_\infty)_e$.

uitdrukking (17) en aan het feit dat er over Bessel-functies zo veel bekend is. Indien het mogelijk zou zijn (24) direct te vinden zonder de gesloten vorm (17) te gebruiken, zou een gelijksoortige benadering voor de rechthoek, waarvoor geen gesloten vorm bestaat, misschien succes kunnen hebben. Zo'n directe methode voor de ellips wordt in de volgende alinea's geschetst. De analoge berekening voor de rechthoek is veel gecompliceerder en wordt behandeld in het Aanhangsel.

De asymptotische impedantie van de ellips kan worden verkregen via eenvoudige fysische overwegingen bij het schema van fig. 4, die natuurlijk equivalent zijn met wiskundige overwegingen bij de corresponderende kettingbreuk (22). Wij verkrijgen een n'degraads benadering van het schema door de $(n+1)$'e parallelle inductie kort te sluiten of de $(n+1)$'e serieweerstand te openen; de hiermee corresponderende impedanties noemen we respectievelijk z_s en z_0. In het eerste geval gaat het netwerk bij hoge frequenties over in n weerstandseenheden in serie; voor de benaderde impedantie geldt dan:

$$z_s = n. \tag{35}$$

In het tweede geval overheerst bij hoge frequenties de impedantie van de laatste inductie de laatste weerstand, die derhalve kan worden verwaarloosd. De laatste inductie vormt dus met de voorgaande een parallelschakeling, en we kunnen nu dezelfde redenering weer opnieuw houden. Ten slotte gaat het netwerk dus over in een parallelschakeling van de eerste n inducties. Aangezien de totale susceptantie $2(1 + 2 + 3 + \ldots + n)$ bedraagt, hetgeen voor grote n ongeveer gelijk is aan n^2, vinden we voor de impedantie:

$$z_0 = k/n^2. \tag{36}$$

Volgens de klassieke netwerktheorie convergeren de ingangsimpedanties z_0 en z_s van een dissipatieve vierpool die aan de uitgang geopend of kortgesloten is, naar een gemeenschappelijke waarde z (de karakteristieke impedantie) als de demping in het netwerk of de lengte van de lijn oneindig wordt. In tegenstelling hiermee lopen (35) en (36) bij toenemende n uiteen. Dit is te wijten aan de essentiële singulariteit in het oneindige van de functie (17) als gevolg van zijn asymptotisch gedrag (24). Hoewel de drie impedanties z_0, z_s en z dus divergeren voor grote $|k|$, kunnen we hopen de asymptotische uitdrukking voor z te verkrijgen door aan z_0 en z_s een gemeenschappelijk asymptotisch gedrag op te leggen. Indien, als eerste poging daartoe, (35) en (36) gecombineerd worden met $z_0 = z_s$, levert dit de betrekking

$$n^3/k = 1, \tag{37}$$

en derhalve voor z_0 en z_s de waarde $k^{1/3}$. Daarmede

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

is de asymptotische uitdrukking (24) teruggevonden, afgezien van het kleine verschil tussen 1 en de correcte waarde (25) voor de coëfficiënt C. Deze discrepantie ontstaat doordat de waarden (35) en (36) van z_s en z_0 berekend zijn door k te laten toenemen bij een vaste waarde van n, zonder dus rekening te houden met de beperking (37), die pas achteraf en als een eerste benadering is gevonden. Dit suggereert dat een iteratief proces tot een verbetering zou kunnen leiden, zowel van de voorwaarde (37) als van de waarde voor de coëfficiënt C in (24) die hieruit volgt. De wiskundige rechtvaardiging van dit proces is gebaseerd op de „potentiometrische ongelijkheden" $z_s < z < z_0$, die geldig zijn voor iedere positieve n en k, en die een gemeenschappelijk asymptotisch gedrag aan de drie impedanties opleggen als wordt voorgeschreven dat z_s/z_0 naar 1 gaat. Wegens de uiteenlopende waarden (35) en (36) van z_0 en z_s voor grote k en vaste n, kan de verhouding z_s/z_0 alleen naar 1 gaan als n en k tegelijk naar oneindig gaan onder een of andere (nog onbekende) beperkende voorwaarde. Omdat deze voorwaarde onbekend is, kan zij slechts worden gevonden via opeenvolgende benaderingen, die telkens tot verbeterde schattingen van de asymptotische uitdrukking voor z leiden, daar de door de potentiometrische ongelijkheden toegelaten speling bij iedere stap afneemt. Hoewel nog niet bewezen is dat dit proces convergeert, is de overgebleven speling na één of twee stappen reeds zeer gering.

De tweede benadering die in de plaats moet komen van (35), (36) en (37) vinden we als volgt. Als de eenheidsstroom wordt geïnjecteerd in het netwerk met de impedantie (35), is de ingangsspanning n, en de spanning in het i'de knooppunt is evenredig met $n - i$. De totale magnetische energie in de parallelinducties is dus

$$\frac{1}{2}\sum_{i=1}^{n} 2i(n-i)^2 \approx \frac{n^4}{12}.$$

Dit moet worden gelijkgesteld aan de energie $v^2/2L$ in een equivalente, met de weerstand (35) parallel geschakelde inductie L. Aangezien v gelijk is aan n, is $L = 6/n^2$ en moet (35) worden vervangen door

$$z_s = \frac{n}{1 + n^3/(6k)}. \tag{38}$$

Op een soortgelijke wijze berekenen we de totale dissipatie in het netwerk van de oorspronkelijke reactieve impedantie (36) en representeren deze als een serieweerstand. Deze blijkt de waarde $8n/15$ te bezitten, zodat (36) moet worden vervangen door

$$z_0 = \frac{k}{n^2}\left(1 + \frac{8n^3}{15k}\right). \tag{39}$$

Het is niet toegestaan (38) en (39) exact aan elkaar gelijk te stellen, want de tweedegraadsvergelijking die dan ontstaat bevat termen in $(n^3/k)^2$, die in (38) en (39) zijn verwaarloosd. Men moet daarom de vergelijking lineariseren in de omgeving van de eerste-ordebenadering (37), door n^3/k in de correctiefactoren te vervangen door 1. Daardoor gaat (38) over in $6n/7$ en (39) in $23k/15n^2$. Door deze uitdrukkingen aan elkaar gelijk te stellen, verkrijgen we voor beide de uitdrukking (24) met $C = 1{,}044\ldots$

Wiskundige afleiding

Wij geven nu een wiskundige formulering van het laterale skineffect in een platte geleider voor de ellips (fig. 1*a*) en de rechthoek (fig. 1*b*) te zamen. Daarbij is $2h$ de dikte, waarbij h in het eerste geval door (9) wordt gegeven en in het tweede geval gelijk is aan de constante b. De lineaire stroomdichtheid is de discontinuïteit van de tangentiële component H_x van het magneetveld. Wegens de symmetrie geldt in een rechtsdraaiend coördinatenstelsel:

$$i = -2H_x\Big|_{y=+0}. \tag{40}$$

Anderzijds is de echte stroomdichtheid (A/cm²):

$$\frac{i}{2h} = \varrho E_z. \tag{41}$$

Ten slotte levert de wet van Lenz:

$$\frac{\partial E_z}{\partial x} = \mu j\omega H_y. \tag{42}$$

Gemakshalve introduceren we de vectorpotentiaal A; deze heeft alleen een z-component. Dan geldt:

$$H_x = \frac{1}{\mu}\frac{\partial A_z}{\partial y}; \qquad H_y = -\frac{1}{\mu}\frac{\partial A_z}{\partial x}, \tag{43}$$

terwijl het spanningsverval ZI langs de geleider gelijk is aan

$$-\frac{\partial V}{\partial z} = E_z + j\omega A_z. \tag{44}$$

In (44) kan E_z door middel van (41), (40) en (43) in A_z worden uitgedrukt; dit levert:

$$ZI = j\omega A_z - \frac{1}{\mu\varrho h}\frac{\partial A_z}{\partial y}\Bigg|_{y=+0}. \tag{45}$$

Anderzijds verkrijgen we door eliminatie van i en E_z uit (40), (41) en (42) de randvoorwaarde van Biot [11]:

$$\frac{\partial}{\partial x}\left(\frac{H_x}{\varrho h}\Bigg|_{y=+0}\right) = -\mu j\omega H_y. \tag{46}$$

Uit de betrekking die ontstaat door eliminatie van de magneetvelden uit (46) met behulp van (43), volgt dat de afgeleide van (45) naar x nul is. De voorwaarde van Biot is dus equivalent met de uitspraak dat de impedantie, berekend met (45), onafhankelijk is van x. De vorm (45) van de randvoorwaarde verdient de voorkeur boven (46), aangezien (45) tevens de impedantie direct levert. Het eigenlijke probleem is nu de vergelijking van Laplace voor A_z op te lossen onder de voorwaarden dat op de geleider (d.w.z. voor $y = 0$, $|x| < a$) vergelijking (45) geldt en dat op grote af-

[11] M. Biot, Ann. Soc. Sci. Brux. 51, 94, 1943.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

stand van de geleider de waarde van A_z gegeven wordt door:

$$A_z = -\frac{I\mu}{2\pi} \ln \frac{\sqrt{(x^2+y^2)}}{D}. \tag{47}$$

Dit is de vectorpotentiaal rondom een draad met stroom I in de oorsprong, omgeven door een cilindrische retourleiding met grote straal D.

We voeren de volgende conforme afbeelding in:

$$x + \mathrm{j}y = a \cos (u + \mathrm{j}v). \tag{48}$$

Daaruit volgt:

$$x = a \cosh v \cos u; \qquad y = a \sinh v \sin u. \tag{49}$$

Onder de beperkingen

$$-\pi \leqslant u \leqslant \pi; \qquad v \geqslant 0$$

is de afbeelding één op één. In het x,y-vlak zijn de krommen van constante v homofocale ellipsen (*fig. 7*); het segment $y = 0$, $|x| \leqslant a$ is de oneindig platte ellips $v = 0$, naar buiten neemt v onbegrensd toe. De krommen van constante u zijn de hyperbolen van fig. 7, maar het segment $v = 0$ is een snede: u is positief in het bovenhalfvlak Re $y > 0$ en negatief in het benedenhalfvlak. Het half-oneindige segment $y = 0$, $x \geqslant a$ correspondeert met $u = 0$, het segment $y = 0$, $x \leqslant -a$ met $u = \pm\pi$.

De uitdrukking voor A_z is van de vorm

$$A_z = \frac{I\mu}{2\pi} \left[\ln \frac{2D}{a} - v + \Sigma A_n \mathrm{e}^{-2nv} \cos 2nu \right]. \tag{50}$$

De eerste twee termen geven de hoofdwaarde (47), want volgens (49) geldt:

$$\frac{\sqrt{(x^2+y^2)}}{D} = \frac{a \cosh v}{D} \approx \frac{a\mathrm{e}^v}{2D}.$$

In (50) is de som met onbepaalde coëfficiënten A_n (waarbij n loopt van 1 tot ∞) de algemene harmonische functie, die in het oneindige nul wordt en de juiste symmetrie heeft.

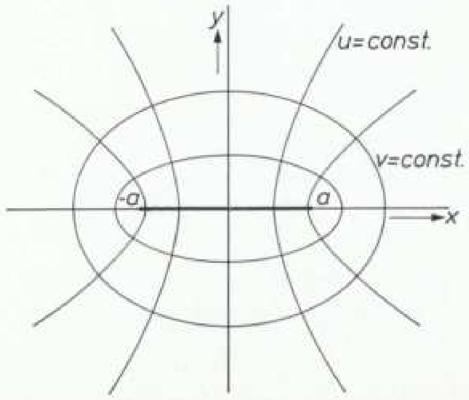


Fig. 7. De conforme afbeelding $x + \mathrm{j}y = a \cos (u + \mathrm{j}v)$.

Voor $y = +0$, dus $v = 0$, $u > 0$, geldt wegens (49):

$$\left.\frac{\partial A_z}{\partial y}\right|_{y=+0} = \frac{1}{a|\sin u|} \left.\frac{\partial A_z}{\partial v}\right|_{v=0},$$

zodat (45) overgaat in:

$$Z = \frac{\mathrm{j}\omega\mu}{2\pi} \left(\ln \frac{2D}{a} + \Sigma A_n \cos 2nu\right) + \\ + \frac{1}{2\pi\varrho ha|\sin u|} (1 + 2\,\Sigma\, n A_n \cos 2nu). \tag{51}$$

De coëfficiënten A_n moeten nu zodanig worden gekozen, dat (51) onafhankelijk van u wordt $(-\pi \leqslant u \leqslant \pi)$.

Voor de elliptische doorsnede gaat de eerste betrekking (49) voor $v = 0$ over in (18), waarmede (9) overgaat in

$$h = b|\sin u|, \tag{52}$$

zodat de noemer in (51) de eenvoudige vorm

$$h|\sin u| = b \sin^2 u = \frac{b}{2}(1 - \cos 2u) \tag{53}$$

krijgt. Met (15), (16) en (12) en de substitutie

$$B_n = -kA_n \tag{54}$$

gaat (51) na vermenigvuldiging met (53) over in:

$$(z + 2\,\Sigma\, B_n \cos 2nu)(1 - \cos 2u) + \\ + \frac{2}{k} \Sigma\, n B_n \cos 2nu = 1. \tag{55}$$

Vervangen we nu in het linkerlid van (55) de produkten van cosinussen nog door cosinussen van sommen en verschillen, dan verkrijgen we een Fourier-reeks; gelijkstellen daarvan aan 1 levert het oneindige systeem

$$\begin{bmatrix} 1 & -1 & 0 & 0 & 0\ldots \\ -1 & 2(1+1/k) & -1 & 0 & 0\ldots \\ 0 & -1 & 2(1+2/k) & -1 & 0\ldots \\ 0 & 0 & -1 & 2(1+3/k) & -1\ldots \\ \ldots & & & & \end{bmatrix} \begin{bmatrix} z \\ B_1 \\ B_2 \\ B_3 \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \end{bmatrix} \tag{56}$$

van lineaire vergelijkingen in z en de onbekende coëfficiënten B_n.

Laten we de eerste twee vergelijkingen van (56) voorlopig even rusten, dan vinden we de recurrente betrekking

$$B_{n-1} + B_{n+1} = 2(1 + n/k)B_n \quad (n = 2, 3, \ldots), \tag{57}$$

die lijkt op de betrekking

$$J_{v-1}(s) + J_{v+1}(s) = \frac{2v}{s} J_v(s) \tag{58}$$

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

voor Bessel-functies. Met $v = n + k$, $s = k$ volgt uit (58) dat

$$B_n = C\, J_{n+k}(k) \qquad (59)$$

met willekeurige C aan (57) voldoet. In feite is (59) de enige oplossing van de tweede-orde-differentievergelijking (57) die voor ons van belang is, omdat bij de andere onafhankelijke oplossing (die een Bessel-functie van de tweede soort bevat) alle coëfficiënten B_n oneindig worden in het geval van gelijkstroom. Omdat (57) geldt vanaf $n = 2$ en dus ook op B_1 betrekking heeft, geldt (59) vanaf $n = 1$. Er blijven dus slechts twee onbekenden over: de factor C in (59) en de impedantie z. Deze worden bepaald door de eerste twee vergelijkingen van (56), die nog niet gebruikt zijn. Lossen we deze op, en maken we gebruik van de betrekking

$$J_v'(s) = \frac{v}{s}\, J_v(s) - J_{v+1}(s), \qquad (60)$$

met $v = s = k$, dan vinden we:

$$C = \frac{1}{J_k'(k)}, \qquad (61)$$

alsmede vergelijking (17).

Wij kunnen nu A_n bepalen met behulp van (54), (59) en (61), vervolgens A_z met behulp van (50), daaruit H_x berekenen met behulp van (43) en ten slotte i met behulp van (40). Het resultaat is verg. (19). We verkrijgen de vorm (20) wanneer we i met behulp van (41) berekenen, waarbij E_z uit (44) volgt met $-\partial V/\partial z = ZI$. Daarmede is het bewijs van de basisresultaten voor de ellips, (17) tot en met (21), geleverd.

De ontwikkeling (22) van (17) volgt uit de recurrente betrekking (57). In het vervangingsschema van fig. 4 komen deze overeen met de wet van Kirchhoff voor de stromen bij de knooppunten. Derhalve is het volgende „elektrische bewijs" equivalent met een wiskundige beschouwing van (57). We beschouwen het laddernetwerk van *fig. 8*, met de serie-admittanties Y_1, Y_2, . . . en de parallel-admittanties Y_a, Y_b, ... We noemen de spanningen tussen de knooppunten en aarde V_i. Wordt de eenheidsstroom geïnjecteerd aan de ingang, dan leidt een elementaire knooppuntsanalyse tot het stelsel lineaire vergelijkingen

$$\begin{bmatrix} Y_1 & -Y_1 & 0 & 0 & \dots \\ -Y_1 & Y_1+Y_a+Y_2 & -Y_2 & 0 & \dots \\ 0 & -Y_2 & Y_2+Y_b+Y_3 & -Y_3 & \dots \\ \dots & & & & \end{bmatrix} \begin{bmatrix} V_0 \\ V_1 \\ V_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix} \qquad (62)$$

voor de spanningen; de oplossing V_0 is de ingangsimpedantie. Vergelijken we (56) met (62), dan zien we dat de oplossing z van (56) de ingangsimpedantie is van het laddernetwerk van fig. 4, waarbij de elementen zijn genormeerd (k is de complexe frequentie). Aangezien de ingangsimpedantie van fig. 4 door (22) wordt gegeven, hebben we indirect de kettingbreukontwikkeling van de functie (17) verkregen.

Voor de rechthoekige doorsnede is h constant in de noemer van (51). Na vermenigvuldiging met de Fourier-reeks

$$|\sin u| = \frac{4}{\pi}\left[\frac{1}{2} - \frac{\cos 2u}{1\times 3} - \frac{\cos 4u}{3\times 5} - \dots\right] \qquad (63)$$

en substitutie van (54), gaat (51) over in:

$$(z + 2\,\Sigma\, B_n \cos 2nu)\left(1 - \frac{2\cos 2u}{1\times 3} - \frac{2\cos 4u}{3\times 5} - \dots\right) + \\ + \frac{2}{k}\,\Sigma\, n\, B_n \cos 2nu = 1. \qquad (64)$$

Het lineaire stelsel vergelijkingen dat uit (64) volgt is analoog met (56), maar de matrix wordt nu door (28) gegeven. Daarmede is (27) bewezen.

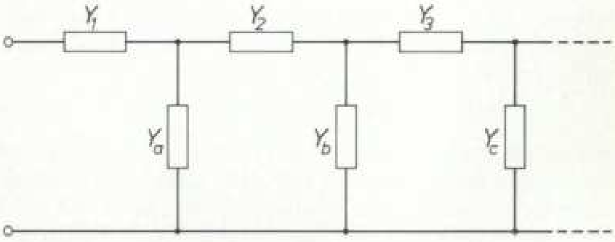


Fig. 8. Een laddernetwerk.

Aanhangsel: Hf.-impedantie van de rechthoekige band

Aan het eind van het hoofdstuk over polen en nulpunten zijn we erin geslaagd een goede benadering te vinden voor de asymptotische impedantie van de ellips zonder de gesloten uitdrukking (17) te gebruiken. Het resultaat was (24), maar in plaats van de juiste waarde 1,088 . . . voor de coëfficiënt C (zie (25)) vonden we in opeenvolgende benaderingen 1 en 1,044 . . . . In dit Aanhangsel passen we dezelfde methode toe op de rechthoek. De analyse is echter veel moeilijker doordat de matrix (28) een volle matrix is, terwijl de matrix in (56) slechts tri-diagonaal was. Daarom werken we nu alleen de eerste benadering uit. Als gevolg daarvan zullen we de nauwkeurigheid van het resultaat niet kunnen schatten. Ook kunnen de numerieke resultaten van Silvester niet als verificatie dienen, daartoe bestrijken zij een te klein frequentiegebied. De betekenis van ons resultaat is dus beperkt. Het is echter het enige thans beschikbare; het kan door verder onderzoek worden verbeterd, en de gebruikte methode is op zichzelf van wiskundig belang.

Evenals bij de ellips is de nu volgende analyse gebaseerd op een vervangingsschema, maar dat kan worden vertaald in zuiver wiskundige termen. Omdat de matrix (28) niet tri-diagonaal is, is het circuit nu geen laddernetwerk maar een algemener RL-netwerk met een oneindig aantal knooppunten. Aangezien echter de termen met k^{-1} alleen op de diagonaal voorkomen, verbin-

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

den de inducties slechts de knooppunten met aarde, de verbindingen van de knooppunten onderling zijn zuiver resistief. Voorts is van iedere rij in (28) de som van de elementen voor $k = \infty$ gelijk aan nul, zoals blijkt uit (63) met $u = 0$. Dit betekent dat bij frequentie oneindig de conductantie tussen ieder knooppunt en aarde nul is, zodat de verbinding tussen knooppunt n en aarde alleen uit de inductie $1/2n$ bestaat, als in fig. 4.

Breekt men het netwerk na n knooppunten af door het openen van alle weerstanden die naar verdere knooppunten leiden, dan is wat overblijft een weerstandsnetwerk met n knooppunten, met een inductie van ieder knooppunt naar aarde, welk netwerk geen uit zuivere weerstanden bestaande verbinding van het eerste knooppunt met aarde bevat. De impedantie tussen knooppunt 0 en aarde is bij hoge frequenties dus oneindig; de hoofdwaarde ervan wordt dus alleen door de inducties bepaald, zodat alle weerstanden kunnen worden kortgesloten. De resulterende impedantie is dus die van een parallelschakeling van de eerste n inducties, en heeft dus evenals bij de ellips de waarde (36).

Bij de complementaire methode van afbreken na n knooppunten worden alle verdere knooppunten verbonden met aarde. De weerstanden die naar verdere knooppunten leiden leveren dan een uit weerstanden bestaande verbinding van knooppunt 0 met aarde, zodat de impedantie bij hoge frequenties zuiver resistief is, evenals in (35); de waarde ervan kan worden bepaald door alle inducties te openen. Dit komt neer op het berekenen van (27) voor $k = \infty$ met een matrix (28) die na orde n is afgebroken. We merken op dat door het afbreken de som van de elementen in iedere rij niet langer nul is, zodat de matrix niet-singulier is. De berekening van de impedantie in deze benadering is equivalent met het oplossen van z uit (64) met

$$k = \infty, \quad B_{n+1} = B_{n+2} = \ldots = 0. \qquad (65)$$

We merken nog op dat het afbreken van de Fourier-reeks met coëfficiënten B_i correspondeert met het kortsluiten van de hogere knooppunten, maar dat de Fourier-reeks (63) die in (64) voorkomt niet afgebroken wordt: deze correspondeert met de aangehouden weerstandsverbindingen met hogere knooppunten. Wegens (63) is de door (65) gemodificeerde vorm van (64):

$$z_s + 2\sum_{i=1}^{n} B_i \cos 2iu = \frac{2}{\pi\,|\sin u|}\,. \qquad (66)$$

In de oorspronkelijke (niet-afgebroken) vorm (64), die ontstaan is uit de randvoorwaarde (45), werden de coëfficiënten B_i en de impedantie z bepaald door z onafhankelijk van u te stellen. Door het afbreken is dit niet meer exact mogelijk; z_s kan slechts in n punten (het aantal onbepaalde coëfficiënten) constant worden gesteld; wegens de symmetrie van (66) kunnen deze alle in het eerste kwadrant worden gekozen. De symmetrie blijft behouden als men n equidistante punten kiest op onderlinge afstand $\pi/2n$ met een afstand $\pi/4n$ tot de uiteinden. Deze interpolatiepunten zijn:

$$u = \frac{\pi}{4n}, \frac{3\pi}{4n}, \frac{5\pi}{4n}, \ldots, \frac{(2n-1)\pi}{4n}.$$

Bij deze klassieke trigonometrische interpolatie is de „gelijkstroomcomponent" z_s van (66) eenvoudig het gemiddelde van de tweede term van (66) in de interpolatiepunten:

$$z_s = \frac{2}{n\pi}\sum_{i=0}^{n-1}\frac{1}{\sin(2i+1)\pi/4n}. \qquad (67)$$

Aangezien de gegeven afleiding van (67) weinig direct is, gaan we na of dezelfde afleiding de waarde (35) levert in het elliptische geval. Vergelijking (55) onder de beperking (65) levert:

$$z_s + 2\sum_{i=1}^{n} B_i \cos 2iu = \frac{1}{1-\cos 2u} = \frac{1}{2\sin^2 u},$$

en in plaats van (67) vinden we:

$$z_s = \frac{1}{2n}\sum_{i=0}^{n-1}\frac{1}{\sin^2(2i+1)\pi/4n}, \qquad (68)$$

hetgeen inderdaad met n identiek is [12].

Voor (67) is geen gesloten uitdrukking beschikbaar, maar voor grote n kan (67) benaderd worden met behulp van de sommatieformule van Euler-MacLaurin, toegepast op de functie

$$\frac{1}{\sin x} - \frac{1}{x},$$

waaruit de singulariteit in $x = 0$ verwijderd is. Op deze wijze vindt men als benadering voor (67):

$$z_s = \frac{4}{\pi^2}\ln\frac{16n\gamma}{\pi}, \qquad (69)$$

waarin γ de constante van Euler is; $\gamma = 1{,}781\ldots$.

Identificatie van (36) met (69) en eliminantie van n levert:

$$k = z\left(\frac{\pi}{16\gamma}\right)^2 e^{\pi^2 z/2}; \qquad (70)$$

dit is een betrekking die z impliciet definieert als functie van k. Door het reële en het imaginaire gedeelte van de logaritme van (70) van elkaar te scheiden, kan men laten zien dat het imaginaire gedeelte van z eindig blijft voor $k = \mathrm{j}\Omega$, terwijl het reële gedeelte r naar oneindig gaat. Uit de modulus van (70) leidt men dan af:

$$\Omega = r\left(\frac{\pi}{16\gamma}\right)^2 e^{\pi^2 r/2}. \qquad (71)$$

De schrijver is dank verschuldigd aan prof. C. Manneback die het idee heeft geopperd de platte ellips te bestuderen in het coördinatenstelsel (48), aan prof. dr. C. J. Bouwkamp en dr. A. C. Vliegenthart voor hulp bij het afleiden van (69), aan C. Wellekens en J. Praet voor de computerresultaten, en aan dr. Y. Genin voor vele vruchtbare discussies.

[12] I. S. Gradshteyn en I. M. Ryzhik, Table of integrals, series, and products, 4e ed., Academic Press, New York 1965, paragraaf 1.382, formule (1) met $x = 0$.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info
11/14/2023

Het Hoofdkantoor van de N.V. Philips' Gloeilampenfabrieken te Eindhoven, zetel van de Raad van Bestuur.

Digitized in Heiloo the Netherlands
Not for sale!

www.freeservicemanuals.info 11/14/2023

Deutsche Luftbild K.G. (Freigabe-Nr. D.L.A. Hbg: 2859/69)

Philips Forschungslaboratorium Hamburg.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Philips techn. T. **32**, 389-402, 1971, no. 10/11/12

Magneto-akoestische effecten in bismut

K. Walther

Inleiding

Bismut is een halfmetaal dat door zijn uitzonderlijke elektronische eigenschappen lange tijd interessant is geweest voor wetenschappelijk onderzoek. In hoofdlijnen kunnen de vaste stoffen in drie klassen verdeeld worden: metalen, halfmetalen en halfgeleiders. De toegestane energietoestanden in vaste stoffen zijn in verscheidene energiebanden gegroepeerd, die onderling gescheiden zijn door verboden energiezones. De bezetting van deze energiebanden wordt beschreven door de Fermi-Dirac-statistiek. Bij lage temperaturen zijn alle toestanden tot aan een zekere energie E_F, de Fermi-energie genaamd, bezet en zijn de toestanden met hogere energie leeg. De beweging van ladingdragers in vaste stoffen kan worden beschreven in termen van een effectieve massa m^*, die door wisselwerking met de atomen van het rooster in het algemeen verschillend is van de massa m_0 van het vrije elektron.

In een *metaal* is het aantal valentie-elektronen van dien aard dat er gedeeltelijk gevulde energiebanden zijn. Doordat er geen verboden energiezone is die binnen een band de bezette van de onbezette toestanden scheidt, is elektrische geleiding zelfs bij lage temperaturen mogelijk. De concentratie van ladingdragers die deelnemen aan de elektrische geleiding is vergelijkbaar met de concentratie van atomen in het rooster en bedraagt ongeveer 10^{22} tot 10^{23} cm^{-3}.

In een *halfgeleider*, waar een eindige verboden energiezone bestaat tussen valentie- en geleidingsband, is de situatie geheel anders. Meestal is bij lage temperaturen de valentieband geheel gevuld en de geleidingsband leeg. Alleen bij hogere temperaturen is elektrische geleiding mogelijk en wel hetzij door het thermisch exciteren van elektronen uit de valentieband naar de geleidingsband (intrinsieke geleiding), hetzij door de halfgeleider te doteren, d.w.z. door een zeker aantal elektronen en gaten in de geleidings- en de valentiebanden te creëren (extrinsieke geleiding). In halfgeleiders is de concentratie van ladingdragers die bijdragen tot de elektrische geleiding (10^{14} tot 10^{19} cm^{-3}) veel kleiner dan de concentratie van atomen in het rooster.

Een *halfmetaal*, zoals bismut, wordt gekarakteriseerd door het feit dat de top van de valentieband bij een iets hogere energie ligt dan de bodem van de geleidingsband (*fig. 1*). De ligging van het Fermi-niveau dat de lege van de bezette toestanden scheidt is zo dat er in de valentieband een kleine concentratie p van zich als gaten in het geleidingsproces gedragende onbezette toestanden aanwezig is, en een corresponderende concentratie van elektronen in de geleidingsband. De ladingdragersconcentratie is veel kleiner dan de concentratie van atomen in het rooster. Elektrische geleiding is bij alle temperaturen mogelijk aangezien er nu geen verboden energiezone tussen bezette en lege toestanden in een band is.

De uitzonderlijke elektronische eigenschappen van bismut die dit halfmetaal tot een interessant object van onderzoek gemaakt hebben, komen vooral tot uiting bij temperaturen T in het gebied van die van vloeibaar helium, waar de relaxatietijden τ, die de verstrooiing van ladingdragers door het rooster beschrijven, erg lang zijn ($\tau \approx 10^{-9}$ s). De effectieve massa's zijn in bismut erg klein ($m^*/m_0 \approx 0{,}01$ voor elektronen), hetgeen resulteert in zeer grote beweeglijkheid van de ladingdragers: $\mu = e\tau/m^*$, waarin e de lading van het elektron is; bij lage temperatuur is μ voor elektronen ca. 10^8 cm^2/Vs. In de ongedoteerde toestand is bismut een „gecompenseerd" halfmetaal met gelijke concentraties van elektronen en gaten: $n = p \approx 3\times10^{17}$ cm^{-3} bij $T = 4{,}2$ K.

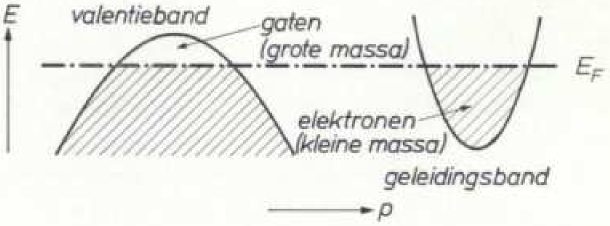


Fig. 1. Energiebandstructuur van een halfmetaal zoals bismut. Schematische voorstelling van de relatie tussen de energie E en impuls p van de ladingdragers voor de valentie- en de geleidingsband. E_F is de Fermi-energie. Beide banden zijn tot aan het Fermi-niveau met elektronen gevuld. De effectieve massa m^* van de gaten in de valentieband is groot, die van elektronen in de geleidingsband is slechts een kleine fractie van de massa m_0 van een vrij elektron.

Deze elektronische eigenschappen leiden tot de, aan het einde van de 19e eeuw ontdekte, ongebruikelijke galvanomagnetische effecten in bismut. De magnetoweerstand in bismut is zeer veel groter dan in andere materialen doordat een transversaal magnetisch veld

Dr. K. Walther, medewerker van Philips Forschungslaboratorium Hamburg GmbH (PFH), Hamburg, Duitsland.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

zowel de elektronen als de gaten naar dezelfde kant van het monster afbuigt, waardoor geen ruimtelading kan worden gevormd.

Dank zij de kleine ladingdragersconcentraties en de kleine effectieve massa's werd bismut het eerste materiaal waarin als functie van de magnetische veldsterkte zg. oscillatorische quantumeffecten werden ontdekt. In 1930 namen Schubnikow en De Haas [1] karakteristieke oscillaties in de magnetoweerstand van bismut waar. De Haas en Van Alphen [2] vonden soortgelijke oscillaties in de magnetische susceptibiliteit. Nadien werden deze effecten toegeschreven aan de gequantiseerde beweging van ladingdragers loodrecht op het magnetisch veld. Nog later werden de oscillatorische quantumeffecten een krachtig hulpmiddel in het onderzoek van de elektronische bandenstructuur van vele vaste stoffen.

De interactie tussen ultrasone golven en ladingdragers in bismut bij aanwezigheid van een magnetisch veld werd het eerst onderzocht door Reneker [3]. Een hetzij longitudinaal hetzij transversaal gepolariseerde akoestische golf veroorzaakt een sinusoïdaal variërende mechanische spanning ε in het kristal en verschuift de energieën van de ladingdragers periodiek met een bedrag δE dat evenredig is met de spanning: $\delta E = C\varepsilon$. De evenredigheidsfactor C wordt de „deformatiepotentiaal" van de desbetreffende energieband genoemd. De situatie in metalen en halfgeleiders met slechts één type ladingdrager is in het algemeen nogal ingewikkeld, doordat door de beweging van de ladingdragers elektrische ruimteladingsvelden ontstaan die de primaire kracht ten gevolge van de deformatiepotentiaal gedeeltelijk afschermen. Deze moeilijkheid heeft men niet in een gecompenseerd halfmetaal zoals bismut, waarin zich geen ruimteladingsveld kan ontwikkelen doordat de akoestische golf elektronen en gaten gelijktijdig verplaatst. Magneto-akoestische onderzoekingen vormen daarom een bruikbare methode om de onafgeschermde deformatiepotentialen van elektronen en gaten in bismut te bepalen [4] [5]. Dit is het eerste onderwerp dat in dit artikel besproken zal worden.

Het tweede in dit artikel te behandelen onderwerp is dat van de magnetodynamische effecten ten gevolge van de aanwezigheid van een constant elektrisch veld waardoor de ladingdragers een driftsnelheid v_d krijgen. De belangstelling voor dit onderwerp werd gestimuleerd door de mogelijkheid van toepassingen, gelijksoortig aan die in een lopendegolfversterker, waarin een microgolf die zich langs een vertragingslijn voortplant, in wisselwerking treedt met een elektronenbundel. Analoog verkrijgt een akoestische golf in een vaste stof energie uit de zich in een bepaalde richting bewegende ladingdragers mits de driftsnelheid in de richting van de akoestische voortplanting groter is dan de geluidssnelheid v_s. Dit betekent dat een ultrasone golf kan worden versterkt door een elektrisch driftveld aan te leggen. Dit effect werd het eerst gedemonstreerd in de piëzo-elektrische halfgeleider CdS door Hutson, McFee en White [6]. Een akoesto-dynamisch effect in bismut werd voor het eerst waargenomen door Esaki [7], die bij een driftsnelheid v_d gelijk aan v_s een knik vond in de stroom-spanningskarakteristiek van een bismutmonster in een transversaal magnetisch veld. Deze knik werd op de volgende wijze geïnterpreteerd [8]. Als aan de conditie voor ultrasone versterking, $v_d > v_s$, voldaan is, genereren de driftende ladingdragers in een breed frequentiegebied ultrasone ruis. Deze statistisch veranderende akoestische golf resulteert in een extra-gelijkstroomcomponent I_{ae} (de akoesto-elektrische stroom genaamd), welke in de stroom-spanningskarakteristiek van het monster tot uitdrukking komt. Wij hebben de ultrasone ruis in bismut direct kunnen waarnemen [9]. De eerste theoretische onderzoekingen die betrekking hadden op ultrasone versterking in halfmetalen werden gepubliceerd door Dumke en Haering [10], en experimentele aanwijzingen voor het optreden van ultrasone versterking in bismut door Toxen en Tansal [11]. Nadien werd de invloed van een elektrisch driftveld op de magneto-akoestische oscillaties in bismut experimenteel bestudeerd [12]. De maximaal bereikbare ultrasone versterking in bismut [13] wordt hoofdzakelijk begrensd door de interactie met de tegelijkertijd gegenereerde ruis.

Bandstructuur en deformatiepotentiaal in bismut

De energiebandstructuur in vaste stoffen beschrijft de relatie tussen de energie E en de impuls p van de ladingdragers, waarbij de wisselwerking met de periodiek in het kristalrooster gerangschikte atomen in de beschouwingen wordt betrokken. Voor een elektron in de vrije ruimte is deze relatie:

$$E = \frac{p^2}{2m_0} = \frac{p_1^2 + p_2^2 + p_3^2}{2m_0}, \tag{1}$$

waarbij p_1, p_2 en p_3 de componenten van de impuls in de drie orthogonale richtingen in de „impulsruimte" zijn. Voor ladingdragers in een vaste stof moet de massa van het vrije elektron m_0 vervangen worden door een effectieve massa die in het algemeen anisotroop is in de impulsruimte (componenten m_1^*, m_2^*, m_3^*), hetgeen dan leidt tot:

$$E = \frac{1}{2}\left(\frac{p_1^2}{m_1^*} + \frac{p_2^2}{m_2^*} + \frac{p_3^2}{m_3^*}\right). \tag{2}$$

In deze vergelijkingen rekent men de energie en de

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

impuls voor elektronen vanaf de onderkant van de geleidingsband en voor gaten vanaf de bovenkant van de valentieband (zie fig. 1). De relatie tussen energie en impuls in verg. (2) leidt voor verschillende richtingen in de impulsruimte tot parabolen (zie *fig. 2*), die in een richting p_1 corresponderend met een grote massa m_1^* zwak gekromd zijn en een sterke kromming vertonen in een met kleine massa m_2^* corresponderende richting p_2. Door de koppeling via de deformatiepotentiaal veranderen de ladingdragers hun energie- en impulstoestanden onder invloed van een geluidsgolf. Zulke overgangen zijn alleen mogelijk in de buurt van de grens tussen bezette en onbezette toestanden, d.w.z. vlak bij de Fermi-energie. Daarom is de voor akoestische effecten interessante grootheid het oppervlak van constante energie $E = E_F$ in de impulsruimte, hetgeen het Fermi-oppervlak genoemd wordt. Uit verg. (2) kan men zien dat dit oppervlak een ellipsoïde is waarvan de lange as evenwijdig is met de met grote massa corresponderende richting p_1 en de korte as evenwijdig met de met kleine massa corresponderende richting p_2 (zie fig. 2). Een energieband van deze vorm wordt „ellipsoïdaal-parabolisch" genoemd.

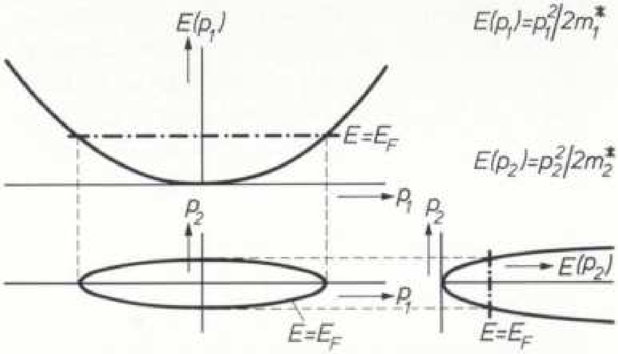


Fig. 2. Ellipsoïdaal-parabolische energieband. De betrekkingen $E(p)$ zijn parabolisch, de oppervlakken E = const. in de impulsruimte zijn ellipsoïden. De effectieve massa m_1^* is groot, de massa m_2^* is klein.

De kristalstructuur van bismut heeft trigonale symmetrie. De ligging van de ellipsoïdale Fermi-oppervlakken van elektronen en gaten wordt ten opzichte van de volgende kristallografische assen beschreven: x = de binaire as, y = de bissectrice, z = de trigonale as (*fig. 3*). Het Fermi-oppervlak van de elektronen bestaat uit drie door de bandindices l = 1, 2, 3 genummerde ellipsoïden die met trigonale symmetrie (120° hoekverschil) rond de z-as gerangschikt zijn. De op de grote massa betrekking hebbende richting van de elektronenellipsoïden is over een kleine hoek (ca. 6°) ten opzichte van de y-as gedraaid. Het Fermi-oppervlak van de gaten bestaat uit één enkele omwentelingsellipsoïde om de z-as, welke genummerd wordt door de bandindex l = 4. Zijn met grote massa corresponderende richting is evenwijdig met de z-as.

Omdat de hoofdassen van de ellipsoïden van constante energie in bismut niet samenvallen met de kristallografische assen, moet verg. (2) voor elke energieband l = 1, . . . 4 in een meer algemene vorm worden geschreven:

$$E = \tfrac{1}{2} \sum_{i,k=1}^{3} \alpha_{ik} p_i p_k . \qquad (3)$$

De anisotropie van de reciproke effectieve massa wordt hier beschreven door coëfficiënten α_{ik} in te voeren, die in de vorm van een matrix gerangschikt kunnen worden. Voor de elektronenellipsoïde l = 1 vindt men:

$$(\alpha_{ik})_{l=1} = \begin{pmatrix} \alpha_1 & 0 & 0 \\ 0 & \alpha_2 & \alpha_4 \\ 0 & \alpha_4 & \alpha_3 \end{pmatrix} . \qquad (4)$$

De component α_4 brengt de draaiing ten opzichte van de y-as in rekening. De elektronenbanden l = 2, 3 worden verkregen door deze ellipsoïde over ± 120° om de trigonale as te draaien.

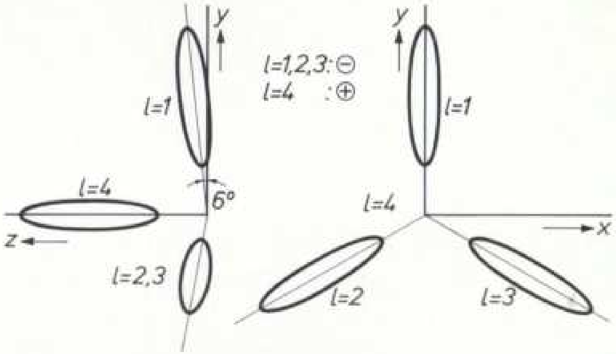


Fig. 3. De ligging van de ellipsoïden van constante energie voor elektronen (l = 1, 2, 3) en gaten (l = 4) in bismut. De kristalstructuur heeft trigonale symmetrie: x is de binaire as, y de bissectrice en z de trigonale as.

[1] L. Schubnikow en W. J. de Haas, Proc. Sec. Sci. Kon. Akad. Wetensch. Amsterdam **33**, 130, 350, 363, 418 en 433, 1930 (Comm. Phys. Lab. Univ. Leiden **19**, nos. 207a, 207c, 207d, 210a en 210b).
[2] W. J. de Haas en P. M. van Alphen, Proc. Sec. Sci. Kon. Akad. Wetensch. Amsterdam **33**, 680 en 1106, 1930, en **35**, 454, 1932 (Comm. Phys. Lab./Kamerlingh Onnes Lab. Univ. Leiden **19** en **20**, nos. 208d, 212a en 220 d).
[3] D. H. Reneker, Phys. Rev. **115**, 303, 1959.
[4] S. Inoue en M. Tsjuji, J. Phys. Soc. Japan **22**, 1191, 1967.
[5] K. Walther, Phys. Rev. **174**, 782, 1968.
[6] A. R. Hutson, J. H. McFee en D. L. White, Phys. Rev. Letters **7**, 237, 1961.
[7] L. Esaki, Phys. Rev. Letters **8**, 4, 1962.
[8] A. R. Hutson, Phys. Rev. Letters **9**, 296, 1962.
[9] K. Walther, Phys. Rev. Letters **15**, 706, 1965.
[10] W. P. Dumke en R. R. Haering, Phys. Rev. **126**, 1974, 1962.
[11] A. M. Toxen en S. Tansal, Phys. Rev. Letters **10**, 481, 1963.
[12] K. Walther, Phys. Rev. Letters **16**, 642, 1966, en Z. Naturf. **21a**, 1443, 1966.
[13] K. Walther, Solid State Comm. **4**, 341, 1966.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

De matrix voor de gatenellipsoïde ($l = 4$) wordt gegeven door:

$$(\alpha_{ik})_{l=4} = \begin{pmatrix} \beta_1 & 0 & 0 \\ 0 & \beta_1 & 0 \\ 0 & 0 & \beta_2 \end{pmatrix}, \qquad (5)$$

waarbij het verschijnen van twee gelijke componenten β_1 daarin de rotatiesymmetrie om de trigonale as beschrijft.

De energieverandering δE van de ladingdragers als functie van de mechanische spanning ε:

$$\delta E = C\varepsilon \qquad (6)$$

(C is de deformatiepotentiaal), levert in het algemeen voor elke energieband $l = 1, \ldots 4$ een anisotrope betrekking op. Deze kan in rekening gebracht worden door een anisotrope deformatiepotentiaal met componenten $(C_{ik})_l$ in te voeren, die dezelfde symmetrie heeft als de reciproke effectieve massa van de desbetreffende band. De componenten kunnen uit de vergelijkingen (4) en (5) worden verkregen door in plaats van α_i en β_i, a_i en b_i te schrijven. De elasticiteitstheorie laat zien dat de mechanische vervorming in een vaste stof eveneens een anisotrope grootheid is, die samenhangt met de ruimtelijke afgeleiden $\partial e_i/\partial x_k$ van de verplaatsingscomponenten e_i naar de coördinaatassen x_k. In een vlakke akoestische golf met hoekfrequentie ω en golfvectorcomponenten q_k is de verplaatsing evenredig met

$$e_i \exp\left[\mathrm{i}\left(\sum_{k=1}^{3} q_k x_k - \omega t\right)\right] \quad \text{en} \quad \partial e_i/\partial x_k \sim e_i q_k .$$

De algemene anisotrope vorm die in de plaats komt van verg. (6) luidt derhalve:

$$\delta E = \sum_{i,k=1}^{3} C_{ik} e_i q_k \equiv (\boldsymbol{e} \cdot \tilde{C} \cdot \boldsymbol{q}) . \qquad (7)$$

Laten wij, als een specifiek voorbeeld voor de interactie, een schuifspanningsgolf (shear wave) beschouwen die zich langs de x-as voortplant. Dan is als gevolg van de transversale polarisatie $q_1 \neq 0$ en $e_1 = 0$. De koppeling wordt beschreven door de elementen C_{21} en C_{31} die nul zijn voor de elektronenband $l = 1$ en de gatenband $l = 4$ (zie verg. (4) en (5)). Daarom koppelt deze trillingswijze alleen met de elektronenbanden $l = 2$ en 3 (zie ook fig. 6).

Om alle componenten C_{ik} van de deformatiepotentiaal voor elektronen en gaten in bismut te meten, is een aantal ultrasone absorptie-experimenten met bekende richtingen van de golfvector $\boldsymbol{q}$ en de polarisatievector $\boldsymbol{e}$ uitgevoerd (gebruik makend van longitudinale en transversale trillingen), welke verscheidene lineaire combinaties van onbekende coëfficiënten C_{ik} opleveren waaruit deze dan kunnen worden berekend. Een moeilijkheid die zich bij dergelijke experimenten voordoet is het feit dat in het algemeen bijdragen van alle vier de banden $l = 1, \ldots 4$ gelijktijdig in de ultrasone verzwakking voorkomen. De toepassing van een quantiserend magnetisch veld is een bruikbare methode om deze verschillende bijdragen te scheiden. Dit zal in het volgende hoofdstuk worden besproken.

Quantumoscillaties van de magneto-akoestische verzwakking in bismut

De aanwezigheid van een magnetisch veld met fluxdichtheid $\boldsymbol{B}$ verandert de relatie tussen energie en impuls van de ladingdragers drastisch. Als de beweging in semiklassieke termen wordt beschreven, zorgt de Lorentz-kracht $e\boldsymbol{v}\times\boldsymbol{B}$, die loodrecht op $\boldsymbol{B}$ staat, ervoor dat de ladingdragers zich gaan bewegen in een vlak loodrecht op het magnetisch veld, terwijl de component p_b van de impuls die evenwijdig is met $\boldsymbol{B}$, niet beïnvloed wordt. In de impulsruimte bewegen de ladingdragers zich langs een elliptische kromme op de Fermi-ellipsoïde $E = E_F$, die loodrecht op $\boldsymbol{B}$ staat (*fig. 4*). Deze baan wordt periodiek doorlopen met een hoekfrequentie die gelijk is aan de cyclotronfrequentie:

$$\omega_c = \frac{e}{m_c} B . \qquad (8)$$

Hierin is m_c de cyclotronmassa van de baan, welke gelijk is aan het meetkundig gemiddelde van de kleinste en de grootste massa die langs de baan optreden. In het algemeen verandert m_c met de richting van $\boldsymbol{B}$ ten opzichte van de Fermi-ellipsoïde. De quantumtheorie laat zien dat de energie van de cyclotronbeweging $\hbar\omega_c$ gequantiseerd is met een baanquantumgetal $n = 0, 1, 2, \ldots$, en dat de energieniveaus van een ladingdrager in een magnetisch veld worden gegeven door

$$E(n,k_b) = (n + \tfrac{1}{2})\hbar\omega_c + \hbar^2 k^2/(2m_b) , \qquad (9)$$

waarin $\hbar$ de constante van Planck gedeeld door 2π is. De tweede term is de energie van de translatie evenwijdig met $\boldsymbol{B}$, waarvoor een effectieve massa m_b geldt. In de golfmechanische beschrijving van de ladingdragers vervangt het golfgetal k_b de impulscomponent $p_b = \hbar k_b$ als variabele. verloop. De door verg. (9) gegeven energieniveaus worden Landau-niveaus genoemd en vertonen een parabolische $E(p_b)$-afhankelijkheid (*fig. 5*). Het energieverschil tussen naburige Landau-niveaus is $\hbar\omega_c$.

Het magneetveld heeft een grote invloed op de verdeling van de energietoestanden in de impulsruimte. Zonder veld zijn de toegestane toestanden uniform verdeeld over het volume van de $\boldsymbol{p}$-ruimte, en de dichtheid $N(E)$ van toestanden met energieën tussen E en

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

$E + \Delta E$ neemt evenredig met $\sqrt{E}$ toe; zie fig. 5. In een magnetisch veld is er in de buurt van de door verg. (9) gegeven gequantiseerde cyclotronbanen een concentratie van toegestane toestanden, en is het aantal toestanden in de rest van de impulsruimte uitgedund. Dit leidt, zoals in fig. 5 te zien is, tot singulariteiten in de toestandsdichtheid nabij de bodem van de Landau-niveaus.

Wij gaan nu het elementaire proces van akoestische absorptie in quantummechanische termen vertalen. Een elektron in de begintoestand ν met energie $E_\nu = E(n,k_{\mathrm{b,o}})$ en golfgetal $k_{\mathrm{b,o}}$ absorbeert een akoestisch quantum (fonon) met energie $\hbar\omega$ en golfgetal q_{b} evenwijdig met $\boldsymbol{B}$, en gaat dan over in een eindtoestand met energie $E_{\nu'} = E(n',k'_{\mathrm{b,o}})$ en golfgetal $k'_{\mathrm{b,o}}$. Behoud van energie en impuls leiden tot:

$$\begin{aligned} E_{\nu'} - E_\nu &= \hbar\omega, \\ k'_{\mathrm{b,o}} &= k_{\mathrm{b,o}} + q_{\mathrm{b}}. \end{aligned} \qquad (10)$$

In een voldoend sterk magneetveld is de frequentie ω van het fonon veel kleiner dan de cyclotronfrequentie ω_c; zo is b.v. $\omega/\omega_c \approx 2 \times 10^{-4}$ voor $f = 60$ MHz en elektronen in bismut met $m_c = 0{,}01 m_0$ bij $B = 1$ kilogauss (kG). Daarom is in de meeste gevallen de energie van een fonon onvoldoende om overgangen van elektronen tussen Landau-niveaus met verschillende n te bewerkstelligen. Laatstgenoemde overgangen worden

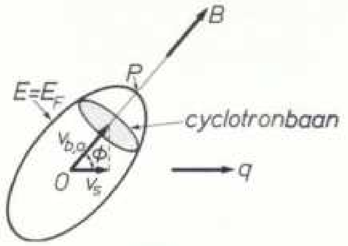


Fig. 4. Een cyclotronbaan van een ladingdrager op de Fermi-ellipsoïde in een vlak (geschaduwd) loodrecht op het magnetische veld $\boldsymbol{B}$. De richting van de geluidsvoortplanting is die van de vector $\boldsymbol{q}$.

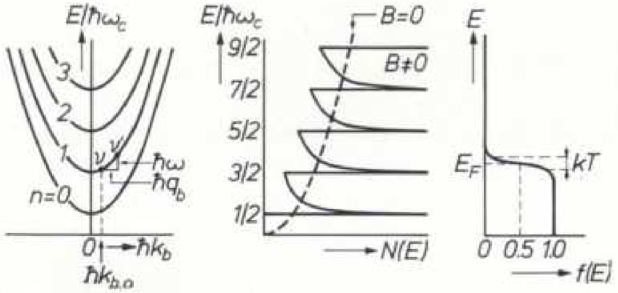


Fig. 5. Landau-niveaus met baanquantumgetal $n = 0, \ldots 3$ in aanwezigheid van een magnetisch veld, de toestandsdichtheid $N(E)$ en de Fermi-Dirac-verdelingsfunctie $f(E)$. De punten ν en ν' representeren de begin- en eindtoestand bij absorptie van een fonon met frequentie ω en golfvectorcomponent q_{b}. Voor deze overgang is $\Delta n = 0$.

waargenomen in cyclotronresonantie-experimenten in het microgolfgebied. Voor ultrasone onderzoekingen liggen de toestanden ν en ν' bij hetzelfde Landau-niveau (zie fig. 5). Als we in de vergelijkingen (9) en (10) de selectieregel $n - n' = 0$ invoeren, komt er:

$$E_{\nu'} - E_\nu = \hbar\omega \approx \frac{\hbar^2 k_{\mathrm{b,o}} q_{\mathrm{b}}}{m_{\mathrm{b}}}, \qquad (11)$$

waarin een kleine term, evenredig met q_{b}^2, verwaarloosd is. Deze vergelijking kan in de vorm

$$v_{\mathrm{s}} = v_{\mathrm{b,o}} \cos \phi \qquad (12)$$

worden herschreven door de snelheidscomponent evenwijdig met het magneetveld $v_{\mathrm{b,o}} = p_{\mathrm{b,o}}/m_{\mathrm{b}} = \hbar k_{\mathrm{b,o}}/m_{\mathrm{b}}$ en de geluidssnelheid $v_{\mathrm{s}} = \omega/q$ in te voeren en $q_{\mathrm{b}} = q \cos \phi$ te schrijven, waarin ϕ de hoek tussen de richtingen van akoestische voortplanting en magnetisch veld is (zie fig. 4). Verg. (12) zegt dat de projectie van de met $\boldsymbol{B}$ evenwijdige snelheidscomponent van de deeltjes op de akoestische voortplantingsrichting gelijk moet zijn aan de geluidssnelheid (de „surf-riding"-conditie voor het absorptieproces). Uit alle mogelijke loodrecht op het magnetisch veld staande baanvlakken selecteert deze voorwaarde voor elke hoek ϕ een effectief vlak. De afstand $v_{\mathrm{b,o}}$ tot het centrum O van de Fermi-ellipsoïde neemt toe met de hoek ϕ en bereikt zijn grootst mogelijke waarde, de Fermi-snelheid v_{F}, bij de kritische hoek $\phi_c = \arccos (v_{\mathrm{s}}/v_{\mathrm{F}})$, waar het baanvlak degenereert tot het punt P op de Fermi-ellipsoïde (zie fig. 4). Voor elektronen in bismut is $v_{\mathrm{s}}/v_{\mathrm{F}}$ ongeveer 10^{-2} tot 10^{-3} en ligt de kritische hoek erg dicht bij $\pi/2$: het verschil tussen $\pi/2$ en ϕ_c bedraagt slechts 0,05° tot 0,5° [2]. Als $\phi > \phi_c$ is, kan aan de conditie van behoud van impuls, verg. (12), niet worden voldaan. In dit gebied heeft Reneker [3] een scherpe daling in de magneto-akoestische verzwakking van bismut waargenomen (het „draaieffect"), waaruit de Fermi-snelheid kon worden berekend. Henrich [14] heeft niet-maximale doorsneden van het Fermi-oppervlak corresponderend met $v_{\mathrm{b,o}} \neq 0$ in de magneto-akoestische verzwakking van bismut gemeten, waarbij ϕ naar $\pi/2$ werd gedraaid. In de meeste gevallen kan voor veldrichtingen die niet loodrecht op v_{s} staan het draaieffect worden verwaarloosd en is $v_{\mathrm{b,o}}/v_{\mathrm{F}}$ buitengewoon klein. Bij de magneto-akoestische absorptie neemt men derhalve baanvlakken waar, die loodrecht op $\boldsymbol{B}$ staan, dicht bij de maximale doorsnede van de Fermi-ellipsoïde.

De bezettingswaarschijnlijkheid van de energietoestanden wordt gegeven door de Fermi-Dirac-functie

$$f(E) = \frac{1}{\exp\,[(E - E_{\mathrm{F}})/kT] + 1}, \qquad (13)$$

[14] V. E. Henrich, Phys. Rev. Letters 26, 891, 1971.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

die één wordt beneden de Fermi-energie E_F en nul daarboven (zie fig. 5). De breedte van het overgangsgebied wordt bepaald door de thermische energie kT, waarin k de constante van Boltzmann is. De netto akoestische absorptie is evenredig met het verschil in de bezettingswaarschijnlijkheden van de elektronentoestanden ν en ν', hetgeen maximaal is bij $E = E_F$ (zie fig. 5) en met behulp van verg. (10) en (13) kan worden benaderd door:

$$f_{\nu'} - f_\nu \approx \left(\frac{\partial f}{\partial E}\right)(E_{\nu'} - E_\nu) \sim \frac{h\omega}{kT}\cosh^{-2}\left(\frac{E_\nu - E_F}{2kT}\right). \qquad (14)$$

In het gebied van quantiserende magnetische velden kan de ultrasone verzwakking $\alpha(B)$ worden berekend [15]:

$$\alpha(B) \sim \frac{\omega}{T}B(\hat{e}\cdot\tilde{C}\cdot\hat{q})^2\frac{m_b}{v_s^2\cos\phi} \times$$
$$\times \sum_n \cosh^{-2}\left(\frac{E(n,k_{b,o}) - E_F}{2kT}\right). \qquad (15)$$

Hierin komt de koppelingsfactor met een deformatiepotentiaal volgens verg. (7) voor, waarin $\hat{e}$ en $\hat{q}$ de eenheidsvectoren in de richtingen van akoestische polarisatie en voortplanting zijn. De factor B in verg. (15) is een consequentie van het feit dat de afmeting van toegestane cyclotronbanen in de impulsruimte groter wordt met toenemend magnetisch veld en een steeds kleiner aantal banen in het volume van de Fermi-ellipsoïde $E = E_F$ past (zie fig. 5). Tot aan het punt waar het laagste Landau-niveau de Fermi-energie E_F passeert, verandert het aantal ladingdragers in de Fermi-ellipsoïde nauwelijks, d.w.z. als het aantal Landau-niveaus onder E_F afneemt met toenemende B, moet de bezetting van deze niveaus toenemen, waarmee de factor B wordt verklaard. De grootheid $m_b/\cos\phi$ in verg. (15) is afkomstig van een integratie over verg. (11), de conditie voor behoud van energie, en kan worden berekend uit de bandstructuurparameters van bismut.

Fysisch kan het absorptieproces als volgt worden beschreven. Er treedt een scherpe piek in de ultrasone verzwakking op als een singulariteit in de toestandsdichtheid het Fermi-niveau passeert (zie fig. 5). De absorptiepieken zijn periodiek in $1/B$, en als het draaieffect wordt verwaarloosd, kan de periode uit de conditie $E(n,k_{b,o}) = E_F$, met behulp van verg. (8) en (9) worden berekend:

$$\Delta\left(\frac{1}{B}\right) = \frac{1}{B_{n+1}} - \frac{1}{B_n} = \frac{e\,h}{m_c E_F} = \frac{e\,h}{S_{max}}. \qquad (16)$$

Hierin is S_{max} de maximale doorsnede van de Fermi-ellipsoïde loodrecht op B in de impulsruimte. De verschillende soorten ladingdragers in bismut (banden $l = 1, \ldots 4$) kunnen worden geïdentificeerd door middel van hun karakteristieke oscillatieperiode. De pieken in de magneto-akoestische verzwakking zijn veel scherper dan de oscillaties die waargenomen worden in het De-Haas-Van-Alphen-effect, doordat het baanvlak bij akoestische absorptie bepaald wordt door de scherpe selectieregel (12), terwijl bij het De-Haas-Van-Alphen-effect veel baanvlakken met verschillende doorsneden en oscillatieperioden tot het effect bijdragen. De meeste oscillatorische bijdragen compenseren elkaar door interferentie en alleen de maximale doorsneden, waarvoor $dS/dk_b = 0$ is, leveren een netto oscillatorisch effect.

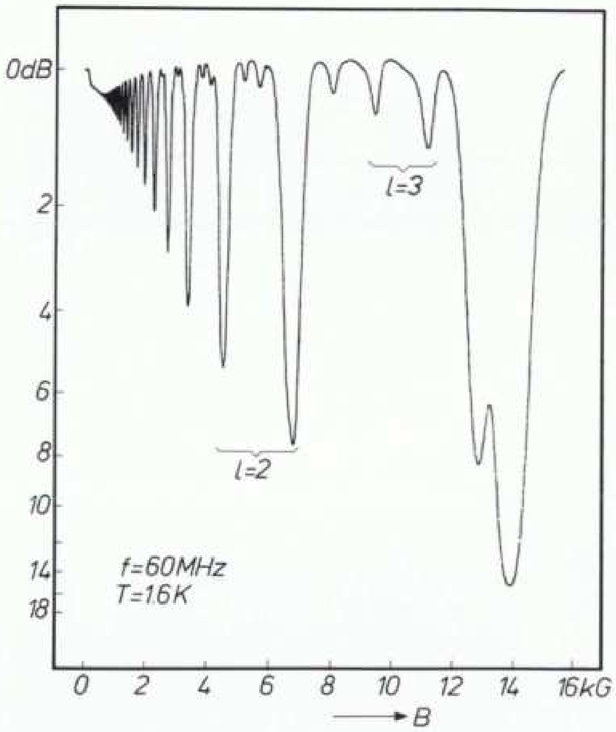


Fig. 6. De amplitude van een doorgelaten signaal voor de zich in de x-richting voortplantende, langzame schuifspanningsgolf bij $f = 60$ MHz en $T = 1{,}6$ K in bismut als functie van het magnetisch veld B. De variaties zijn „quantumoscillaties" van de elektronenbanden $l = 2$ en 3.

Fig. 6 toont gemeten quantumoscillaties in de magneto-akoestische verzwakking voor de langzame schuifspanningsgolf met q evenwijdig met de x-as in bismut, waarin de bijdragen afkomstig van de elektronbanden $l = 2$ en 3 duidelijk gescheiden te zien zijn. De dubbele absorptiepiek bij 13-14 kG is een gevolg van de invloed van de spin van het elektron. In een magnetisch veld splitst elk Landau-niveau zich in twee subniveaus (met spin „omhoog" en spin „omlaag"). De spinsplitsing zullen wij echter niet verder bespreken.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Bepaling van de deformatiepotentialen in bismut

Verscheidene onderzoekers hebben experimentele resultaten betreffende de magneto-akoestische quantumoscillaties in bismut gepubliceerd [3] [16-20]. In de meeste gevallen richtte de belangstelling zich daarbij op het bepalen van cyclotronmassa's, maximale doorsneden van het Fermi-oppervlak en spinsplitsingsfactoren uit de oscillatieperiodes $\Delta(1/B)$. De deformatiepotentialen voor elektronen en gaten in bismut kunnen bepaald worden uit de maxima $\alpha_p(B)$ van de magneto-akoestische verzwakking [4] [5], welke optreden bij $\cosh^{-2}(\ldots) = 1$ in verg. (15).

Dit vereist een zorgvuldige beschouwing van de invloed van botsingen, die in het algemeen de amplitude verkleinen en de breedte van de absorptiepieken vergroten. Er zijn talloze theoretische artikelen over dit onderwerp gepubliceerd [15] [21] [22] waarin wordt aangetoond dat de invloed van strooiing afhangt van het produkt $q_b l_b$, waarin $q_b = q \cos\phi$ en l_b de gemiddelde vrije weglengte van de ladingdragers in de richting van het magnetische veld is. In dit artikel worden de botsingen in aanmerking genomen door de gemeten verzwakking te vermenigvuldigen met een strooiingscorrectiefactor $K(q_b l_b) \geqslant 1$, welke één is voor zeer grote waarden van $q_b l_b$ en toeneemt met afnemende $q_b l_b$. Hoe K van $q_b l_b$ afhangt kan experimenteel bepaald worden.

Om de deformatiepotentialen uit de gemeten maximale verzwakking $\alpha_p(B)$ te berekenen wordt een gereduceerde verzwakking gedefinieerd:

$$\frac{\alpha_p'(B)}{B} = K\frac{\alpha_p(B)}{B}\cdot\frac{v_s^2\cos\phi}{m_b}. \qquad (17)$$

Uit verg. (15) kunnen lineaire combinaties van de diverse componenten van de deformatiepotentiaal worden verkregen door de verhoudingen van de gereduceerde verzwakkingen bij constante ω en T te berekenen voor twee gevallen r en s, die in ultrasone trillingswijze, bandindex of veldrichting ϕ verschillen:

$$\pm\sqrt{\frac{(\alpha_p'(B)/B)_r}{(\alpha_p'(B)/B)_s}} = \frac{(\hat{e}\cdot\tilde{C}\cdot\hat{q})_r}{(\hat{e}\cdot\tilde{C}\cdot\hat{q})_s}. \qquad (18)$$

Door de in verg. (18) voorkomende wortel kan het teken van de deformatiepotentiaal niet uit magneto-akoestische metingen worden afgeleid. Inoue en Tsuji [4] hebben dit teken kunnen vinden door informatie uit magneto-akoestische quantumoscillaties te combineren met piëzogalvanomagnetische metingen in bismut als functie van de statische vervorming uitgevoerd door Jain en Jaggi [23]. De hoekafhankelijkheid van de ultrasone verzwakking in verg. (15) wordt gegeven door

$$\frac{\alpha_p(B)}{B} \sim \frac{m_b}{K(q_b l_b)\cos\phi}, \qquad (19)$$

waarin de gemiddelde vrije weglengte in de richting van het magnetisch veld wordt berekend uit de snelheidscomponent v_b:

$$l_b = \tau v_b = \tau\sqrt{\frac{2E_F}{m_b}},$$

$$q_b l_b = q\cos\phi\cdot\tau\sqrt{\frac{2E_F}{m_b}}. \qquad (20)$$

In bismutmonsters die niet zeer zuiver zijn (verhouding van de elektrische weerstanden $\varrho(300\text{ K})/\varrho(4{,}2\text{ K})$ kleiner dan 200) is de anisotropie van de relaxatietijd τ klein vergeleken met de anisotropie van de effectieve massa [24] (zie echter [25]). Daardoor is, in de met grote massa corresponderende richting van grootste verzwakking (zie verg. 19) de strooiingscorrectiefactor K groot, doordat $q_b l_b$ (zie verg. 20) klein is. In de met kleine massa corresponderende richting wordt de verzwakking klein en kan de strooiingscorrectie worden verwaarloosd; $K \approx 1$, doordat $q_b l_b$ groot is.

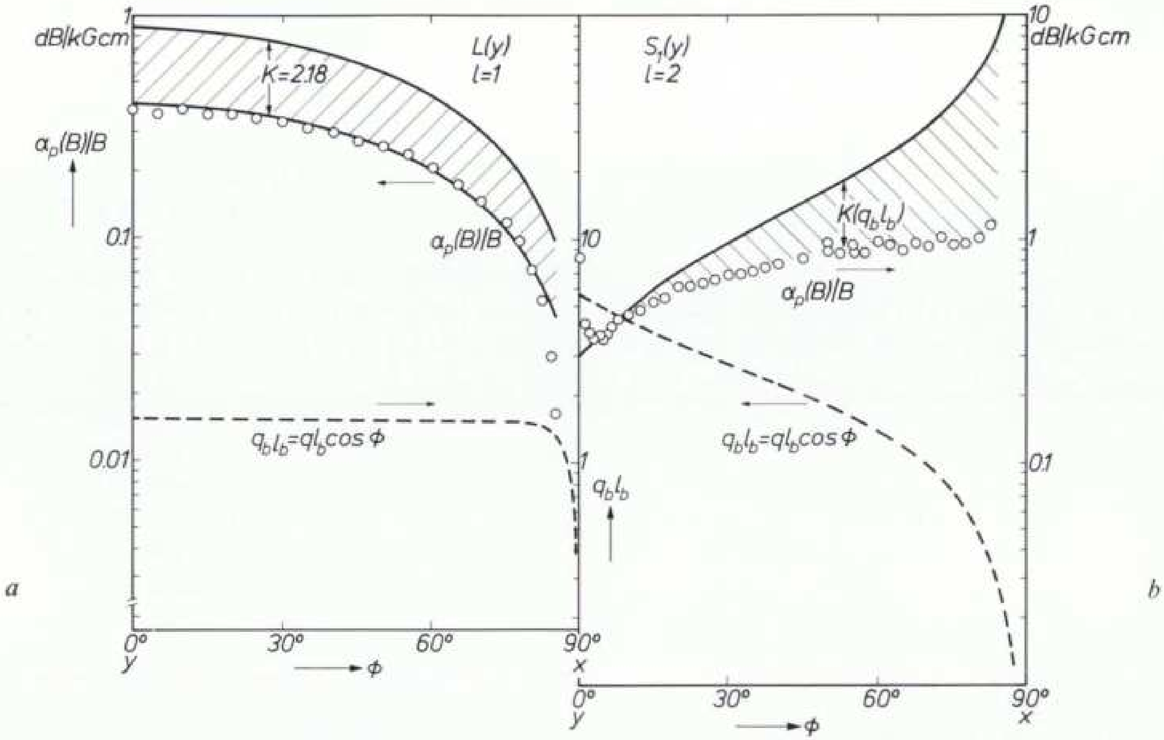
Dit gedrag zal worden toegelicht voor twee specifieke gevallen in bismut. De *figuren 7a* en *7b* laten de hoekafhankelijkheid zien van de maximumverzwakking $\alpha_p(B)/B$ voor respectievelijk de longitudinale golf en de schuifspanningsgolf waarvoor $\hat{e}$ evenwijdig is met de x-as, welke zich voortplanten in de y-richting. De cirkels zijn de experimentele punten bij $T = 1{,}6$ K en $f = 60$ MHz, de getrokken lijnen geven de berekende, met $m_b/\cos\phi$ evenredige hoekafhankelijkheid weer, welke geldig zou zijn als $K = 1$ is. De streeplijnen zijn de grootheden $q_b l_b$, berekend met verg. (20) met behulp van een gemiddelde waarde van de isotrope relaxatietijd $\tau = 1{,}9\times10^{-10}$ s [24] en een Fermi-energie $E_F = 17{,}7$ meV voor elektronen. In fig. 7*b* neemt de koppeling met de elektronenband $l = 2$ toe, en nemen $\cos\phi$ en de gemiddelde vrije weglengte l_b af als de hoek ϕ toeneemt. De resulterende, zeer geprononceerde afneming van $q_b l_b$ correspondeert met een toeneming van de correctiefactor K, die men kan afleiden uit het verschil tussen de experimentele punten en de getrokken theoretische kromme (gearceerd), waarbij $K = 1$ als de twee

[15] V. L. Gurevich, V. G. Skobov en Y. A. Firsov, Zh. eksper. teor. Fiz. **40**, 786, 1961.
[16] A. P. Korolyuk, Fiz. tverd. Tela **5**, 3323, 1963, en Zh. eksper. teor. Fiz. **51**, 697, 1966.
[17] A. M. Toxen en S. Tansal, Phys. Rev. **137**, A211, 1965.
[18] S. Mase, Y. Fujimori en H. Mori, J. Phys. Soc. Japan **21**, 1744, 1966.
[19] Y. Sawada, E. Burstein en L. Testardi, J. Phys. Soc. Japan **21**, Suppl., 760, 1966.
[20] T. Sakai, Y. Matsumoto en S. Mase, J. Phys. Soc. Japan **27**, 862, 1969.
[21] S. H. Liu en A. M. Toxen, Phys. Rev. **138**, A487, 1965.
[22] Een goed overzicht geeft Y. Shapira in: Physical acoustics, red. W. P. Mason, vol. V, Academic Press, New York 1968, blz. 1.
[23] A. L. Jain en R. Jaggi, Phys. Rev. **135**, A708, 1964.
[24] R. N. Zitter, Phys. Rev. **127**, 1471, 1962.
[25] R. Hartman, Phys. Rev. **181**, 1070, 1969.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

krommen elkaar raken bij $\phi = 0°$. Op deze wijze kan de functie $K(q_b l_b)$ experimenteel worden bepaald. In fig. 7*a* neemt de koppeling met de elektronenband $l = 1$ af met toenemende ϕ. De corresponderende toeneming van de vrije weglengte l_b wordt gecompenseerd door de afneming van cos ϕ, hetgeen tot $\phi = 80°$ tot een constante waarde van $q_b l_b$ leidt, nl. 1,55. Door vergelijking met fig. 7*b* wordt een constante strooiingscorrectie-neaire betrekkingen tussen de deformatiepotentialen a_1, a_2 en a_4 voor elektronen worden verkregen. Deze worden in een coördinatensysteem met assen a_1/a_2 en a_4/a_2 in *fig. 8* als rechte lijnen getekend. De twee getallen bij elke lijn verwijzen naar de verschillende combinaties van ultrasone trillingswijze en bandindex en worden in *fig. 9* verklaard. Voor de gemiddelde coördinaten van het gebied waarin de lijnen elkaar krui-



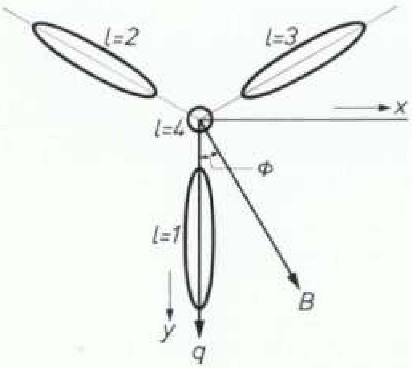


Fig. 7. De verzwakking $\alpha_p(B)/B$ van de longitudinale golf $L(y)$ (elektronenband $l = 1$) en de schuifspanningsgolf $S_1(y)$ (e // x, elektronenband $l = 2$) in bismut, beide zich voortplantend in de y-richting, als functie van de oriëntatie ϕ van het magnetisch veld. De grootheid $q_b l_b$ werd uit verg. (20) berekend.

factor $K = 2{,}18$ verkregen, hetgeen de goede overeenstemming tussen de hoekafhankelijkheid van de theoretische kromme en de experimentele punten verklaart: om de gecorrigeerde verzwakking te vinden moet de experimentele kromme over een constante afstand op de logaritmische schaal verschoven worden.

Door de verhoudingen te berekenen van gereduceerde verzwakkingen voor verscheidene zich in de x- en y-richting voortplantende ultrasone golven, kunnen, met behulp van de vergelijkingen (4), (7), (15) en (18), li-sen vindt men de waarden $a_1/a_2 = -0{,}37$ en $a_4/a_2 = +0{,}25$. Met behulp van deze waarden vindt men uit metingen waarbij de akoestische voortplanting plaatsvindt in de z-richting en langs de bissectrice van de hoek tussen y en z voor de verhouding a_3/a_2 de waarde $-0{,}29$. Tot zover hebben we dus de deformatiepotentialen bepaald in verhouding tot a_2, waarvan de waarde kan worden verkregen door aanpassing van de absolute waarde van de gemeten verzwakking aan die volgens (15). Het resultaat is $a_2 = \pm\, 5{,}9$ eV. De anisotropie van de

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

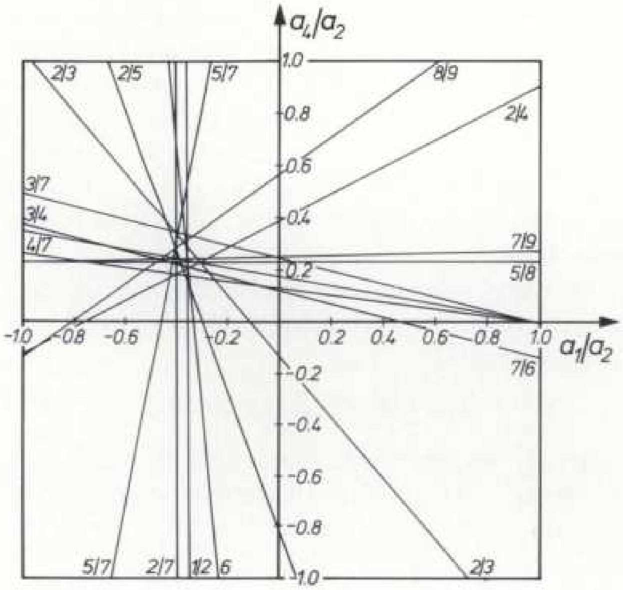


Fig. 8. Bepaling van de verhoudingen a_1/a_2 en a_4/a_2 met behulp van de lineaire betrekkingen tussen de componenten van de deformatiepotentiaal volgens verg. (18). De indices r/s bij elke lijn hebben betrekking op een combinatie van getallen welke in fig. 9 wordt uitgelegd. De lijnen snijden elkaar in een gebied rond $a_1/a_2 = -0{,}37$ en $a_4/a_2 = +0{,}25$.

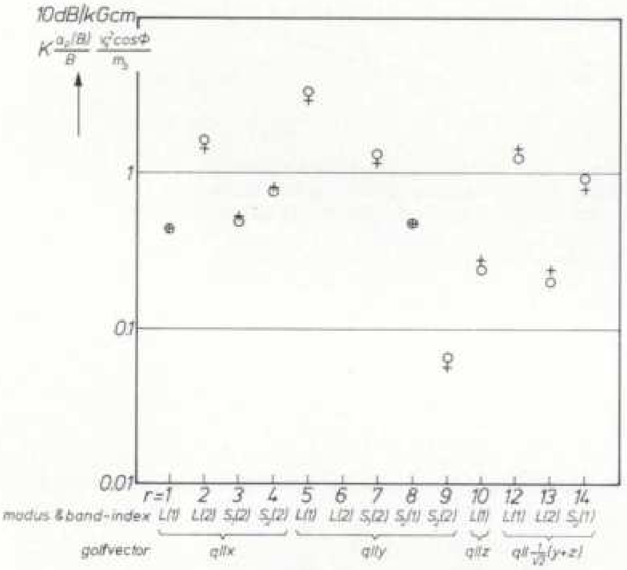


Fig. 9. De gereduceerde ultrasone verzwakking $\alpha_p'(B)/B = K\alpha_p(B)v_s^2(\cos\phi)/Bm_b$ in bismut voor 14 verschillende combinaties van ultrasone trillingswijze en bandindex. o gemeten waarden, + theoretische aanpassing volgens verg. (15), met $a_2 = \pm 5{,}9$ eV.

gemeten magneto-akoestische verzwakking voor de 14 gevallen uit fig. 9 kan met een nauwkeurigheid van 14% theoretisch worden beschreven met de vier deformatiepotentiaalcomponenten voor elektronen. Zo ook kan de deformatiepotentiaal voor gaten bepaald worden: $b_2/b_1 = -1{,}03$ en $b_2 = \pm 1{,}2$ eV. De ultrasone methode verschaft ons een zeer directe manier om de deformatiepotentiaal op vrij veel manieren te meten. De onbepaaldheid in teken, welke inherent is aan deze methode, kan worden opgelost door combinatie met piëzogalvanomagnetische metingen [23]. Op die manier hebben Inoue en Tsuji [4] laten zien dat zowel het teken van a_2 als dat van b_2 negatief is.

De invloed van het elektrische driftveld: het „knik"-effect en de ultrasone ruis in bismut

In de boven gegeven bespreking van de wisselwerking tussen akoestische golven en ladingdragers in bismut hebben we ons beperkt tot het geval dat de driftsnelheid $\boldsymbol{v}_d$ gelijk nul is. Om akoestodynamische effecten te bestuderen in de aanwezigheid van een intermitterend elektrisch driftveld $\boldsymbol{E}$, heeft het voordelen om te werken met een sterk, loodrecht op $\boldsymbol{E}$ staand magnetisch veld $\boldsymbol{B}$ [7], zoals in *fig. 10* geschetst is. Door de grote magnetoweerstand van bismut bij temperaturen van vloeibaar helium zijn de stroomdichtheden die nodig zijn om driftsnelheden groter dan de geluidssnelheid te realiseren, bij deze temperaturen een factor 10^6 kleiner dan in het geval zonder uitwendig magnetisch veld. De oorsprong van de hoge magnetoweerstand van bismut lichten we toe aan de hand van het vectordiagram in *fig. 11*: een sterk magnetisch veld draait de stroomvectoren van elektronen en gaten in tegengestelde richting weg van de richting van het elektrische veld. De draaihoeken, $\Theta_{H,e}$ en $\Theta_{H,h}$, de zg. Hall-hoeken, zijn bijna 90°, aangezien de verhouding van de stroomcomponenten evenwijdig met en loodrecht op $\boldsymbol{E}$ bij $B = 10$ kG en ladingdragerbeweeglijkheden μ van 10^7-10^8 cm²/Vs gegeven wordt door $I_{\parallel}/I_{\perp} = (\mu B)^{-1} \approx \approx 10^{-3}$-$10^{-4}$. Doordat de stroomvectoren voor elektronen en gaten in tegengestelde richting gedraaid worden, is de bovengrens van de afbuighoek, die in unipo-

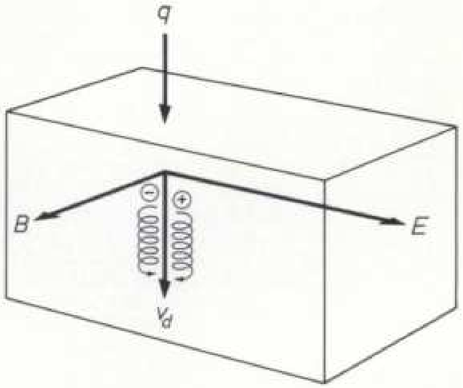


Fig. 10. De beweging van elektronen en gaten in bismut in gekruiste elektrische en magnetisch velden. De gemiddelde driftsnelheid $\boldsymbol{v}_d$ is gelijk aan $\boldsymbol{E}\times\boldsymbol{B}/B^2$. De richting van de geluidsvoortplanting is weer met $\boldsymbol{q}$ aangeduid.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

laire halfgeleiders ten gevolge van het transversale Hall-veld optreedt, in halfmetalen vrijwel afwezig.

De beweging van de driftende ladingdragers moet in de „echte ruimte" worden beschreven (zie fig. 10). In aanwezigheid van gekruiste elektrische en magnetische velden voeren elektronen en gaten cyclotronbewegingen uit met tegengestelde rotatierichting, waarbij (in het vrije-ladingdragersmodel) het centrum van de cyclotronbaan zich verplaatst met een gemiddelde gelijkstroomdriftsnelheid $\boldsymbol{v}_d = \boldsymbol{E}\times\boldsymbol{B}/B^2$, die loodrecht op $\boldsymbol{E}$ en $\boldsymbol{B}$ staat. De anisotropie van de effectieve massa modificeert deze uitdrukking, en Eckstein [26] heeft aangetoond dat voor bepaalde richtingen van $\boldsymbol{E}$ en $\boldsymbol{B}$ zeer hoge driftsnelheden (v_d = 7-8 E/B) kunnen worden verkregen. Om bij B = 10 kG een driftsnelheid E/B te verkrijgen die vergelijkbaar is met de geluidssnelheid

is Ψ de fasehoek tussen E_s en n_s. De deeltjesconcentratie n_s bereikt een maximum als de deeltjes met de geluidssnelheid lopen en neemt af als het snelheidsverschil tussen de deeltjes en de akoestische golf toeneemt. Een ultrasone golf genereert een gelijkstroomcomponent, de „akoesto-elektrische" stroom $I_{ae} \sim |n_s|\,|E_s| \cos\Psi$, die evenredig is met het kwadraat van de mechanische spanning.

De energiebalans tussen de driftende ladingdragers en de geluidsgolf wordt bepaald door de gemiddelde interactiekracht

$$\overline{F}_s = \langle F_s(x)\, n_s(x)\rangle / \langle |n_s(x)| \rangle, \qquad (21)$$

waarin $\langle\ \rangle$ de gemiddelde waarde over één akoestische golflengte is. Afhankelijk van de projectie van de driftsnelheid op de richting van de akoestische voortplan-

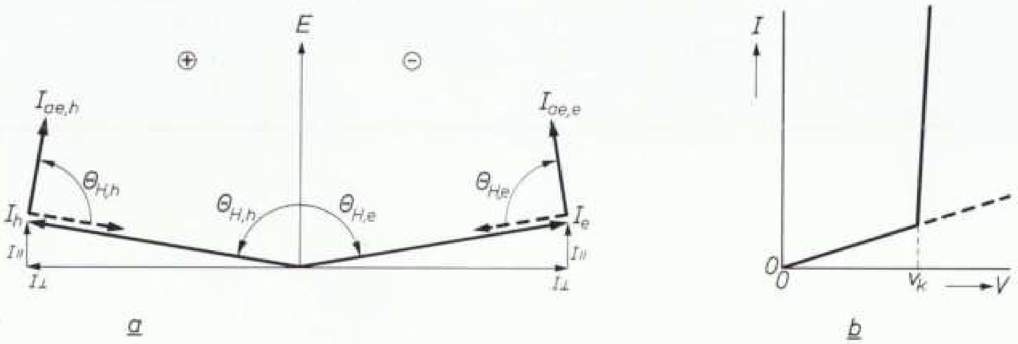


Fig. 11. *a)* Vectordiagram dat de „ohmse" stromen $\boldsymbol{I}$ en de akoesto-elektrische stroomcomponenten I_{ae} voor elektronen (index e) en gaten (index h) in aanwezigheid van een magnetisch veld loodrecht op het vlak van tekening laat zien. De pijl $\boldsymbol{E}$ geeft de richting van het elektrische veld aan. De Hall-hoeken Θ_H zijn bijna 90°. *b)* Knikeffect in bismut. Bij toenemende spanning V wordt boven een zekere waarde V_k de toeneming van de totale stroom veel geprononceerder doordat er dan tevens een toenemende akoesto-elektrische stroomcomponent is. De streeplijn representeert het gedrag vóór het ontstaan van de akoesto-elektrische stroom.

($\approx 10^5$ cm/s) heeft men een elektrisch veld E van 10 V/cm nodig.

Door de interactie via de deformatiepotentiaal (zie 7) genereert een zich b.v. in de x-richting voortplantende akoestische golf een sinusoïdaal variërende potentiaal $\Phi_s(x,t)$ met een bijbehorende veldsterkte

$$E_s(x,t) = \frac{1}{e} F_s = -\frac{1}{e}\frac{\partial \Phi_s}{\partial x}$$

welke gevoeld wordt door de ladingdragers (zie *fig. 12*). Het grootste deel van de ladingdragers loopt met de driftsnelheid $\boldsymbol{v}_d$. Onder invloed van het elektrische veld

$$E_s = |E_s| \exp\,[\mathrm{i}(qx - \omega t)]$$

voert de akoestische golf echter, met de geluidssnelheid, een zekere concentratie

$$n_s = |n_s| \exp\,[\mathrm{i}(qx - \omega t - \Psi)]$$

van ladingdragers met zich mee; in deze uitdrukking

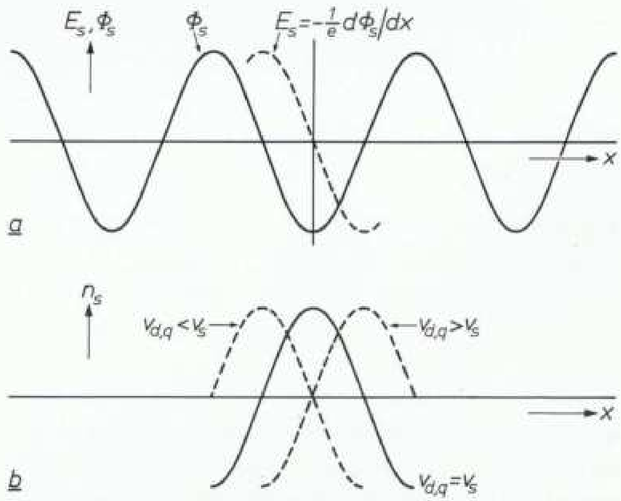


Fig. 12. *a)* De door de geluidsgolf veroorzaakte elektrische potentiaal $\Phi_s(x,t)$ en elektrische veldsterkte $E_s(x,t)$ uitgezet als functie van x. *b)* De ladingdragersconcentratie $n_s(x,t)$ als functie van x voor verschillende driftsnelheden. Als $v_{d,q}$ gelijk is aan de geluidssnelheid v_s is er geen wisselwerking. Als $v_{d,q} > v_s$ is wordt de geluidsgolf versterkt, als $v_{d,q} < v_s$ is wordt de geluidsgolf verzwakt.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

ting ($v_{d,q} = \boldsymbol{v}_d \cdot \hat{\boldsymbol{q}}$), moeten er drie gevallen besproken worden (zie fig. 12):

1) Geen interactie als $v_{d,q} = v_s$: De maxima van de deeltjesconcentratie n_s vallen samen met de minima van de potentiaal Φ_s en de gemiddelde interactiekracht verdwijnt.

2) Ultrasone versterking als $v_{d,q} > v_s$: De maxima in de deeltjesconcentratie n_s zijn ten opzichte van de potentiaalminima verschoven, zodat er een gemiddelde interactiekracht tegengesteld aan de driftbeweging bestaat, welke de deeltjes vertraagt. De golf ontvangt dus energie van de deeltjes en wordt versterkt.

3) Ultrasone verzwakking als $v_{d,q} < v_s$: In vergelijking met het voorgaande geval is de gemiddelde interactiekracht van teken omgekeerd; de ladingdragers worden versneld door het elektrische veld E_s en krijgen deze energie van de akoestische golf, welke wordt verzwakt.

Als aan de voorwaarde voor ultrasone versterking $v_{d,q} > v_s$ is voldaan, genereren de driftende deeltjes in een breed frequentiegebied ultrasone ruis, óók als er van buiten af géén geluidsgolf tot het kristal wordt toegelaten. Deze ultrasone ruis wordt uitgezonden in de richting waarin de deeltjes driften en brengt een akoesto-elektrische gelijkstroomcomponent I_{ae} voort. Esaki [7] heeft aan een Bi-monster bij $T = 2$ K in een sterk transversaal magneetveld bij een spanning V_k waarbij de driftsnelheid de geluidssnelheid begon te overschrijden, een knik waargenomen in de stroom-spanningskarakteristiek. De verhouding van de differentiële weerstanden beneden en boven de knikspanning was groter dan 50. De knik kan worden verklaard met behulp van het vectordiagram van fig. 11*a*: als gevolg van het bestaan van ultrasone ruis boven V_k worden akoesto-elektrische stromen $I_{ae,e}$ en $I_{ae,h}$ van elektronen en gaten opgewekt, in eerste instantie in richtingen die tegengesteld zijn aan die van de „ohmse" stromen I_e en I_h, en vervolgens over de Hall-hoeken $\Theta_{H,e}$ en $\Theta_{H,h}$ gedraaid, waardoor een sterke toeneming van de stroom in de richting van het elektrische veld ontstaat. Wij [9] hebben het met het knikeffect samenhangende optreden van ultrasone ruis waargenomen door gebruik te maken van een langs de *y*-as gesneden kwartstransducent die aan een loodrecht op de *x*-richting staand oppervlak van een bismutmonster was bevestigd. In *fig. 13a* en 13*b* ziet men de ultrasone ruis in het frequentiegebied rond 60 MHz (bovenste spoor), de driftstroom (middelste spoor) en de driftspanning (onderste spoor) bij $T = 1{,}8$ K en $B = 16$ kG. In fig. 13*a* zijn de richtingen van het elektrische en het magnetische veld zo gekozen dat de driftsnelheid $\boldsymbol{v}_d = \boldsymbol{E} \times \boldsymbol{B}/B^2$ naar de kwartstransducent wijst, in fig. 13*b* is de richting van het magnetische veld omgedraaid en wijst de driftsnelheid van de transducent af. De akoesto-elektrische stroom komt tot stand met een incubatietijd van ongeveer 4 µs (middelste spoor), daarmee corresponderend wordt een afneming van de driftsnelheid waargenomen (onderste spoor), en begint de ultrasone ruis 5 µs na het begin van de driftpuls. Als we het veld omkeren wordt de ultrasone ruis 7 µs later gedetecteerd en is zijn amplitude met 20 dB gereduceerd; de gevoeligheid van de oscilloscoop was hier 10 maal zo groot (fig. 13*b*). Dit gedrag laat zien dat de ultrasone ruis in de richting van de ladingdragerdriftsnelheid wordt geëxciteerd: in fig. 13*b* loopt de ruis weg van de transducent en bereikt deze pas weer na gereflecteerd te zijn aan het monsteroppervlak loodrecht op *x*. De extra-vertraging correspondeert bij benadering met de breedte van het monster gedeeld door de geluidssnel-

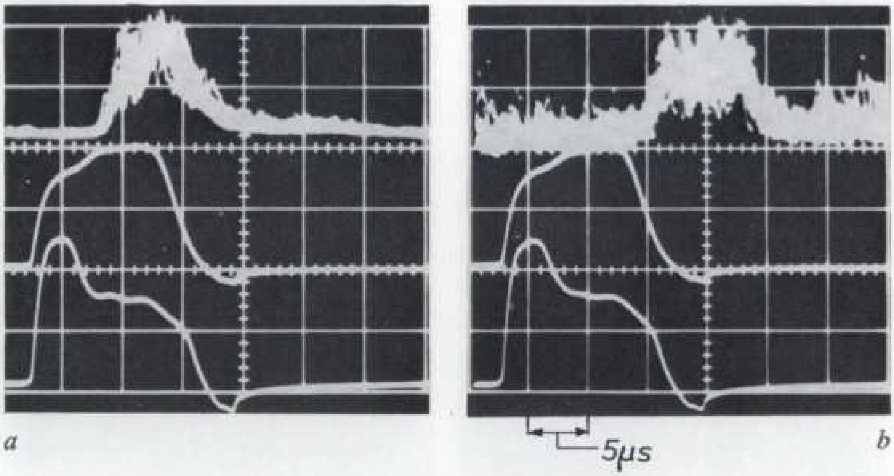


Fig. 13. De bovenste sporen tonen het aangroeien van ultrasone ruis (frequentiegebied rond 60 MHz) langs de *x*-richting in een bismutmonster, $B = 16$ kG, $T = 1{,}8$ K. De middelste sporen tonen de driftstroom (kalibratie 50 A per hoofdschaaldeel) en de onderste sporen de driftspanning (kalibratie 20 V per hoofdschaaldeel). *a*) Driftsnelheid naar de transducent gericht. *b*) Driftsnelheid omgekeerd.

[26] S. G. Eckstein, Physics Letters **13**, 30, 1964.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

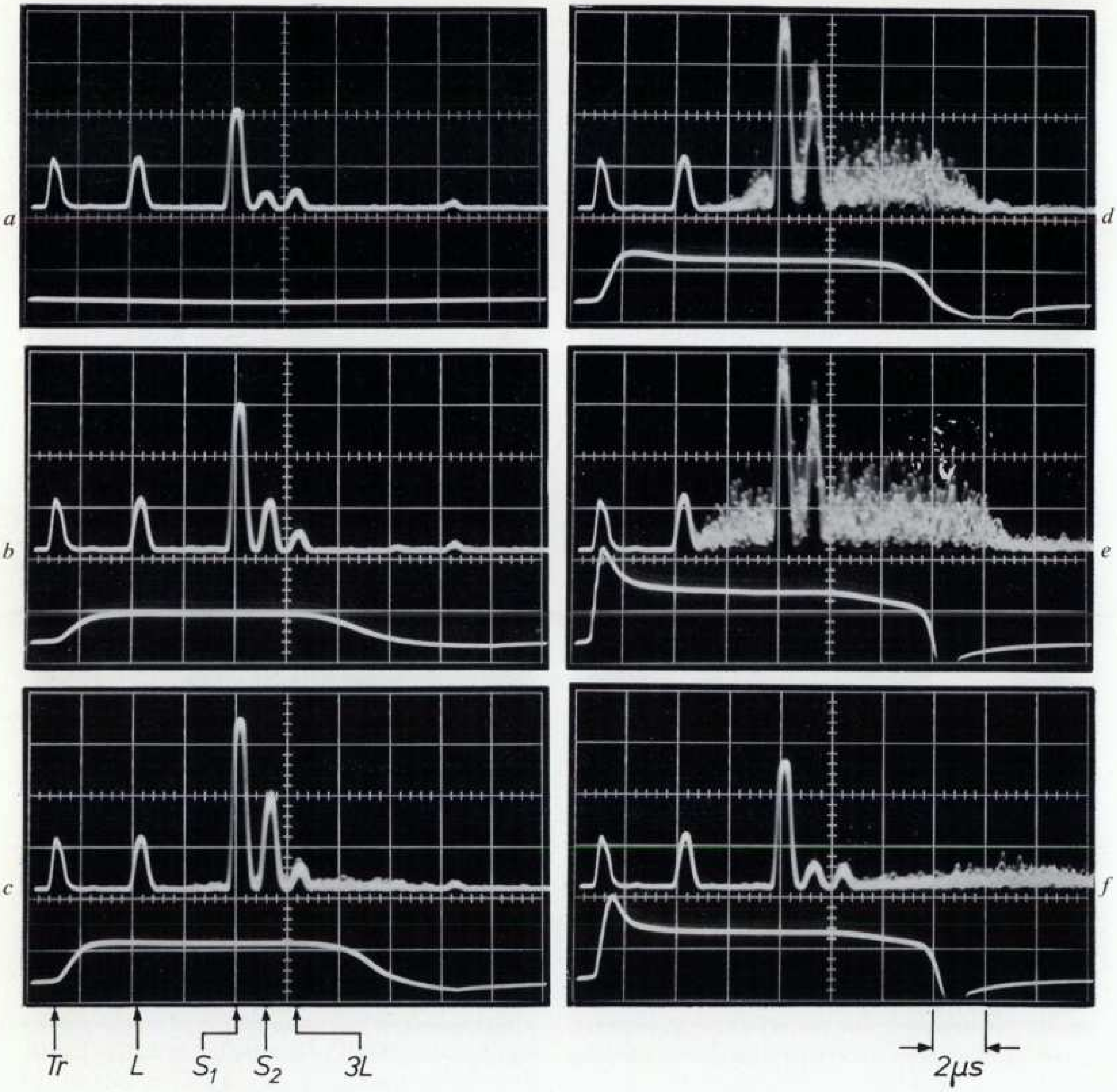


Fig. 14. Ultrasone signaalamplitudes en ultrasone ruis bij f = 60 MHz (bovenste sporen) en driftspanning (onderste sporen) in bismut geregistreerd als functie van de tijd bij B = 13 kG, T = 4,2 K. *a*)-*e*) Driftsnelheid evenwijdig met de geluidsvoortplanting. *f*) Driftsnelheid omgekeerd. De symbolen en de details aangaande de krommen worden in de tekst besproken.

heid van de zich in de x-richting voortplantende langzame schuifspanningsgolf ($v_s = 0{,}89 \times 10^5$ cm/s). De ultrasone ruis werd ook waargenomen in een breed frequentiegebied (tot aan 220 MHz). Door gebruik te maken van driftpulsen met zeer scherpe aanloopflanken heeft Yamada [27] de stroom-spanningskarakteristiek van een bismutmonster gemeten in een aan het aangroeien van de ultrasone ruis en de akoesto-elektrische stroom voorafgaand tijdsinterval; zelfs voor spanningen ver boven het knikpunt, bij voorbeeld $V = 5\,V_k$ (zie fig. 11*b*), vond hij een zuiver „ohms" gedrag.

Ultrasone versterking en geïnverteerde quantumoscillaties in bismut

In dit hoofdstuk bespreken we de invloed van een elektrische driftveld op de verzwakking van een ultrasone golf. Volgens Toxen en Tansal [11] neemt de amplitude van een schuifspanningsgolf van 15 MHz in bismut met 14 dB/cm toe als er een driftveld wordt aangelegd. *In fig. 14* ziet men de amplitudes van de pulsen voor drie ultrasone trillingswijzen (bovenste sporen) met f = 60 MHz en bij T = 4,2 K in bismut als functie van de amplitude van de elektrische driftpuls (onderste sporen) [12]. De eerste puls Tr in de

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

bovenste sporen correspondeert met de elektrische overspraak van de puls van de zender; de met L, S_1 en S_2 aangeduide ultrasone pulsen hebben betrekking respectievelijk op de longitudinale trillingswijze en op de snelle en langzame schuifspanningsgolven. De puls $3L$ is afkomstig van het signaal van de longitudinale trillingswijze die de weg driemaal doorlopen heeft. Er werd gewerkt met een magnetisch veld B van 13 kG, welke waarde, net als in fig. 6, correspondeert met een maximum in de magneto-akoestische verzwakking. In fig. 14*a-e* bewegen de ladingsdragers zich in dezelfde richting als het geluid. De ultrasone amplitude neemt toe met toenemend driftveld. De grootste amplitudeveranderingen, tot 27 dB/cm toe, worden waargenomen in de langzame schuifspanningsgolf S_2. Voorbij het knikpunt wordt ultrasone ruis (zie fig. 14*d* en 14*e*) gegenereerd. In fig. 14*f* is de driftbeweging van de ladingdragers tegengesteld aan de signaalvoortplanting, hetgeen duidelijk te zien is uit het vertraagd aankomen van de ruispuls, en veroorzaakt het driftveld slechts kleine veranderingen in de signaalamplitude. Voor signaal- en ruisvoortplanting in dezelfde richting neemt de signaalamplitude, als de ruisamplitude meetbare waarden bereikt, niet verder toe (fig. 14*d* en 14*e*). Anderzijds voorspelt de theorie van ultrasone versterking voorbij de knikspanning een voortgaande toeneming van de signaalamplitude bij toenemend driftveld. De experimenteel waargenomen bovengrens van de ultrasone versterking wordt kennelijk veroorzaakt door nietlineaire wisselwerking tussen de ultrasone ruis en het ultrasone signaal. De statistische elektrische potentiaal ten gevolge van de ruis, die een grote amplitude heeft, concurreert met de zwakke sinusoïdale modulatie van de elektrische potentiaal door het signaal (zie fig. 12). Dit effect is welbekend van de piëzo-elektrische halfgeleider CdS [28], waarin de incubatietijd tussen de aanloopflank van de driftpuls en het begin van de ultrasone ruis zeer aanzienlijk kan zijn. Door het tijdsinterval van de signaalpuls ten opzichte van de ruispuls te variëren, heeft McFee [28] aangetoond dat de signaalversterking van schuifspanningsgolven van 45 MHz in CdS door de aanwezigheid van ultrasone ruis gereduceerd kan worden tot 23 dB/cm. In het geval van bismut hangt de amplitude van de gegenereerde ultrasone ruis sterk af van de richting van het magnetische veld. Uitgebreide onderzoekingen hebben aangetoond dat bij minimale ruisgeneratie maximale signaalversterking optreedt en omgekeerd.

De invloed van een elektrisch driftveld op de magneto-akoestische quantumoscillaties in bismut ziet men in *fig. 15*; hierin is de amplitude van de langzame schuifspanningsgolf S_2 uitgezet als functie van het magnetisch veld. In het geval van kromme *a*, die betrekking heeft op een frequentie van 60 MHz, is gemeten zonder dat er een elektrisch driftveld aanwezig was. Men vindt dan soortgelijke quantumoscillaties in de magneto-akoestische verzwakking als die in fig. 6. Kromme *b* heeft betrekking op een frequentie van 188 MHz en een meting waarbij een elektrisch driftveld boven het knikpunt aanwezig is. In vergelijking met kromme *a* zijn, als aan de condities voor versterking is

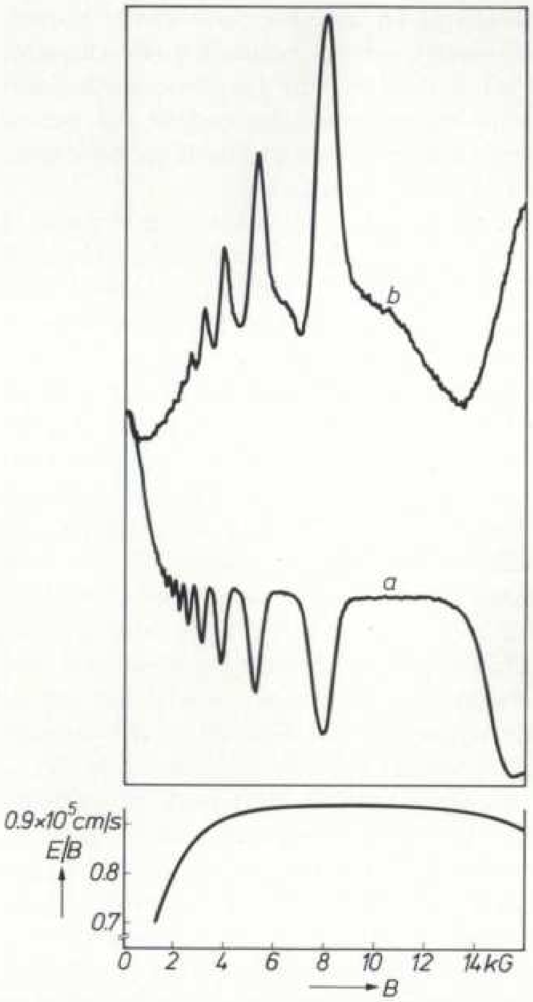


Fig. 15. De amplitude van een ultrasoon signaal voor een langzame schuifspanningsgolf (*boven*) en de driftsnelheid E/B (*beneden*) in bismut als functie van het magnetisch veld bij T = 4,2 K. Kromme *a*: quantumoscillaties bij f = 60 MHz zonder driftveld. Kromme *b*: „geïnverteerde" quantumoscillaties bij f = 188 MHz met driftveld.

voldaan, de quantumoscillaties geïnverteerd: als een singulariteit in de toestandsdichtheid het Fermi-niveau passeert (zie fig. 5), wordt zonder driftveld een maximum in de ultrasone verzwakking waargenomen (kromme *a*), en, bij hetzelfde magnetisch veld, mèt elektrisch driftveld E een maximum in de ultrasone versterking (kromme *b*). Laatstgenoemde kromme werd opgenomen onder condities van optimale signaal-

[27] T. Yamada, J. Phys. Soc. Japan **20**, 1647, 1965.
[28] J. H. McFee, J. appl. Phys. **34**, 1548, 1963.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

versterking boven het knikpunt bij een driftstroom die vrijwel onafhankelijk van het magnetisch veld was. Bij een constante driftstroom voorspelt de theorie het volgende: voor zwakke magnetische velden (zwak knikeffect) is $E \sim B^2$, d.w.z. de driftsnelheid $v_d \approx E/B \sim B$; voor sterke magnetische velden (sterk knikeffect) is $E \approx E_k \sim B$, d.w.z. de driftsnelheid $v_d \approx E/B \approx$ const. [7]. De in het onderste deel van fig. 15 uitgezette driftsnelheid voor kromme *b* is tussen 4 en 16 kG vrijwel onafhankelijk van het magnetisch veld en neemt onder 4 kG af als *B* afneemt. De geïnverteerde quantumoscillatie onder versterkingscondities ziet men in fig. 15 dan ook afgeschilderd als functie van het magnetisch veld bij vrijwel constante v_d.

De beste resultaten voor ultrasone versterking werden verkregen [13] door de langzame schuifspanningsgolf met $q // x$ te gebruiken; het elektrisch veld werd daarbij langs de *z*-richting aangelegd, het magnetisch veld was evenwijdig met de richting van de bissectrice van de hoek tussen de $-y$- en de $+z$-as. *Fig. 16* toont de verandering in ultrasone verzwakking $\alpha - \alpha_0$, ten opzichte van de verzwakking α_0 bij $B = 0$, bij verscheidene frequenties tussen 61 en 267 MHz als functie van de driftsnelheid bij $B = 2{,}15$ kG. De richtingen van akoestische voortplanting en deeltjesdrift zijn voor positieve waarden van de driftsnelheid hetzelfde. Negatieve waarden van v_d betekenen dat de richting van de driftsnelheid omgekeerd is door het magnetisch veld om te draaien. De frequentieafhankelijkheid van de maximale versterking met driftveld en de maximale verzwakking zonder driftveld zijn uitgezet in *fig. 17*. Bij een frequentie van 267 MHz werd een maximale versterking $(\alpha - \alpha_0)_{ampl}$ van $-53{,}5$ dB/cm gemeten, en door extrapolatie blijkt dat de verzwakking zonder driftveld $(\alpha - \alpha_0)_{att}$ gelijk is aan 93,5 dB/cm, zodat het driftveld een totale verandering in verzwakking van ongeveer 147 dB/cm veroorzaakt. Fig. 17 toont ook de geëxtrapoleerde frequentieafhankelijkheid van de verzwakking α_0 zoals gegeven door Reneker [3]; men ziet dat de gemeten waarden van de ultrasone versterking netto geen akoestische winst voor het Bi-kristal opleveren. In het piëzo-elektrische CdS is de situatie anders. Hier kan door ultrasone versterking „nettowinst" worden verkregen [6]. In het geval van bismut wordt de ultrasone versterking bevorderd door de volgende factoren:

De geluidssnelheid van de langzame schuifspanningsgolf is klein: $v_s = 0{,}89 \times 10^5$ cm/s.

De massa-anisotropiefactor in de driftsnelheid [26] is groot voor de elektronenband $l = 2$, waarvoor $v_{d,q} = 1{,}7\ E/B$.

De knik in de driftstroomkromme is bij zwakke magnetische velden erg zwak. Daardoor kan een grote driftsnelheid worden verkregen: $(v_{d,q})_{max}/v_s \approx 2{,}5$.

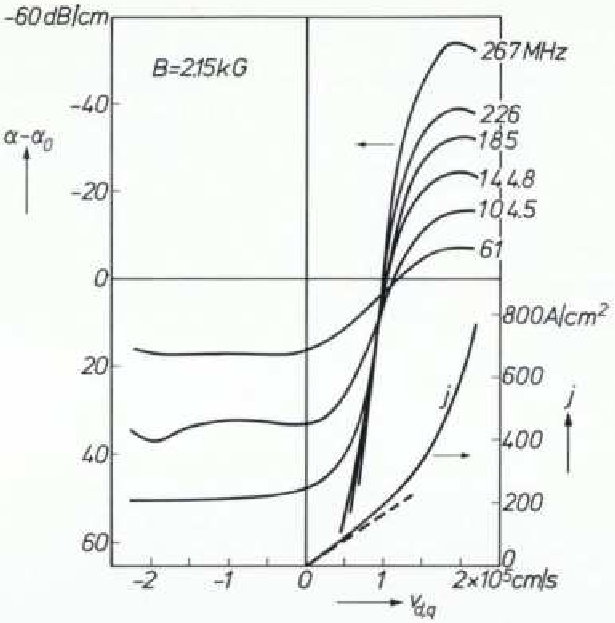


Fig. 16. De verzwakking $\alpha - \alpha_0$ van de langzame schuifspanningsgolf ($q//x$) en de driftstroomdichtheid *j* in bismut als functie van de driftsnelheid $v_{d,q} = 1{,}7\ E/B$ bij $B = 2{,}15$ kG en $T = 4{,}2$ K.

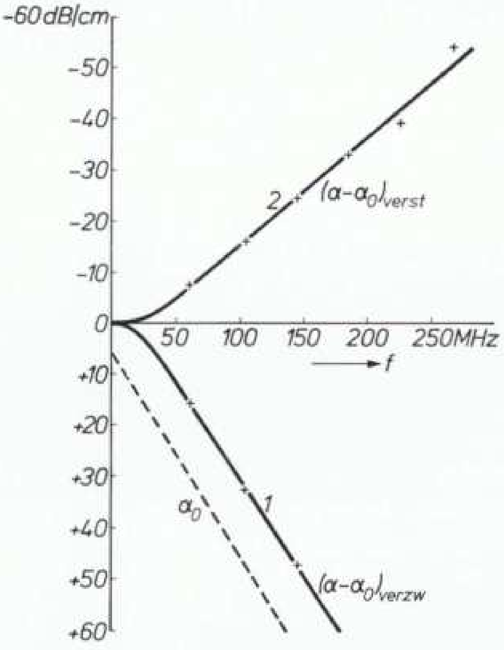


Fig. 17. De maximale veranderingen in de verzwakking zonder driftveld (kromme *1*) en de versterking met driftveld (kromme *2*) voor de langzame schuifspanningsgolf ($q // x$) in bismut als functie van de frequentie *f* bij $B = 2{,}15$ kG en $T = 4{,}2$ K. De streeplijn toont de verzwakking α_0 bij $T = 78$ K volgens Reneker [3].

De opwekking van ultrasone ruis is bij zwakke magnetische velden gering en de incubatietijd voor het ontstaan van ruis is lang [27]. De vermindering van de ultrasone versterking door niet-lineaire wisselwerking met ruis is daardoor klein.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Wervels

J. Volger

Wervels komen in de natuur veel voor: in vloeistoffen, in de beweging van sterren, in supergeleiders, enz. In dit overzichtsartikel worden enige analogieën onder de aandacht van de lezer gebracht. In het eerste hoofdstuk worden enige resultaten van de vectoranalyse gerecapituleerd, te zamen met enige grondbeginselen van de elektriciteitsleer. Daarna worden wervels in een vloeistof, dislocaties in een kristal, fluxdraden in supergeleiders van de tweede soort en wervels in superfluïde helium behandeld. Zowel algemeen bekende aspecten als resultaten van recent onderzoek komen hierbij ter sprake.

Vaak wordt een fysische situatie beschreven door een vectorveld: in ieder punt van de beschouwde ruimte is een vector $\boldsymbol{P}(x,y,z)$ gegeven, bij voorbeeld de snelheid $\boldsymbol{v}(x,y,z)$ van een stromende vloeistof. Er zijn twee operatoren die, toegepast op een vectorveld, belangrijke gegevens verschaffen over de aard van de ruimtelijke variatie van de vector, te weten de divergentie (div) en de rotatie (rot):

$$\operatorname{div} \boldsymbol{P} = \frac{\partial P_x}{\partial x} + \frac{\partial P_y}{\partial y} + \frac{\partial P_z}{\partial z}, \tag{1}$$

$$\operatorname{rot} \boldsymbol{P} = \begin{vmatrix} \boldsymbol{1}_x & \boldsymbol{1}_y & \boldsymbol{1}_z \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ P_x & P_y & P_z \end{vmatrix}. \tag{2}$$

Volgens Gauss geldt:

$$\oint\!\!\oint_S \boldsymbol{P} \cdot \mathrm{d}\boldsymbol{S} = \iiint_V \operatorname{div} \boldsymbol{P}\, \mathrm{d}V. \tag{3}$$

In het geval van een stromingsveld van een onsamendrukbare vloeistof zegt deze stelling dat in punten waar $\operatorname{div} \boldsymbol{v} \neq 0$ is, vloeistof te voorschijn komt — als uit een bron —, of verdwijnt — als in een put.

Ook de rotatie wordt in het hydrodynamische beeld aanschouwelijk. Volgens Stokes is

$$\oint_l \boldsymbol{P}\, \mathrm{d}\boldsymbol{l} = \oint\!\!\oint_S \operatorname{rot} \boldsymbol{P}\, \mathrm{d}\boldsymbol{S}, \tag{4}$$

d.w.z. in het vectorveld wordt de circulatie, dat is de lijnintegraal van de tangentiële component, genomen langs een gesloten contour, gegeven door de totale hoeveelheid rotatie die door de contour wordt omvat. Als ergens de vector rot $\boldsymbol{P}$ ongelijk nul is, kan deze vector, als we hem volgen in de richting waarin hij wijst, niet verdwijnen, want de divergentie van een rotatie is altijd nul: div rot $\boldsymbol{P} \equiv 0$. Vaak is er een cilindertje te onderscheiden waarin rot $\boldsymbol{P} \neq 0$ is, terwijl daarbuiten rot $\boldsymbol{P} = 0$ is. We zullen zo'n cilindertje een wervelas noemen. Een wervelas eindigt op het oppervlak van de vectorruimte — b.v. de wand van het vat waarin zich de stromende vloeistof bevindt — of is in zichzelf gesloten (ringvormige wervel).

Het is mogelijk een vectorveld geheel te bepalen uit de gegeven divergentie en rotatie, d.w.z. uit de bronnen en de wervels. Een door *bronnen* bepaald vectorveld is bij voorbeeld het elektrische veld van een stelsel van vaste elektrische ladingen e_i. De elektrische veldsterkte kan verkregen worden uit een elektrische potentiaal ϕ. Voor deze scalaire veldfunctie geldt:

$$\phi(x,y,z) = \sum \frac{e_i}{r_i(x,y,z)}, \tag{5}$$

waarin $r_i(x,y,z)$ de afstand is van het punt (x,y,z) tot de puntlading e_i. Uit (5) kan men, door toepassing van de operatie „gradiënt", de veldsterkte $\boldsymbol{E}$ afleiden:

$$\boldsymbol{E} = -\operatorname{grad} \phi = -\frac{\partial \phi}{\partial x}\boldsymbol{1}_x - \frac{\partial \phi}{\partial y}\boldsymbol{1}_y - \frac{\partial \phi}{\partial z}\boldsymbol{1}_z. \tag{6}$$

Een voorbeeld van een door *wervels* bepaald vectorveld is het magnetische veld van een draad waarin een elektrische stroom met dichtheid $\boldsymbol{i}$ vloeit. Het magneetveld $\boldsymbol{H}$ moet voldoen aan de Maxwell-vergelijking rot $\boldsymbol{H} = 4\pi\boldsymbol{i}/c$ en wordt verkregen uit een vectorpotentiaal $\boldsymbol{A}$. Voor dit vectorveld geldt:

$$\boldsymbol{A}(x,y,z) = \int \frac{\boldsymbol{i}\,\mathrm{d}V}{c\, r_v(x,y,z)}, \tag{7}$$

waarin $r_v(x,y,z)$ de afstand is tussen het punt (x,y,z) en het volume-element dV. Men verkrijgt $\boldsymbol{H}$ door op (7) de operatie „rotatie", die wij reeds kennen, toe te passen:

$$\boldsymbol{H} = \operatorname{rot} \boldsymbol{A}. \tag{8}$$

Vaak verwondert men zich wanneer men voor het eerst met deze vectorpotentiaal kennis maakt [1]. De

Prof. dr. J. Volger, wetenschappelijk adviseur van Philips' Natuurkundig Laboratorium, Eindhoven.

[1] H. B. G. Casimir, Theorie der electriciteit, Het Kompas, Antwerpen/De Spieghel, Amsterdam, 1936.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

schrijver herinnert zich met name zijn verbazing toen hij narekende welke gedaante $\boldsymbol{H}$ en $\boldsymbol{A}$ hebben in de eenvoudige gevallen van een stroomvoerende oneindig lange draad en van een oneindig lange solenoïde.

Bij een *oneindig lange draad* waardoor een stroom J vloeit (*fig. 1*), is het veld $\boldsymbol{H}$ concentrisch rondom deze draad verdeeld. De sterkte bedraagt $2J/rc$. Verder geldt:

$$\boldsymbol{H} = \text{rot}\,\boldsymbol{A}\,;\ \boldsymbol{A} = A_z \boldsymbol{I}_z\,;\ A_z = \frac{-2J}{c} \ln r + [c]\,. \qquad (9)$$

Het leek begrijpelijk dat $\boldsymbol{H}$ met toenemende afstand van de draad steeds kleiner zou worden, maar verontrustend dat $\boldsymbol{A}$ tot oneindig zou toenemen, zij het ook logaritmisch. De vraag rees: welke fysische betekenis zou deze vectorpotentiaal hebben?

Bij een *oneindig lange solenoïde* die zo is ingericht dat er bij de cilindermantel een tegenstroom vloeit die de totale stroom in de asrichting nul maakt, is buiten de spoel de magnetische veldsterkte nul (*fig. 2*). De vectorpotentiaal is concentrisch rondom de as; voor de sterkte geldt:

$$A = 2\phi/r \text{ buiten}$$
en $\qquad (10)$
$$A = 2\pi r H_0 \text{ binnen,}$$

waarbij H_0 het magneetveld binnen de spoel is en ϕ de magnetische flux van de solenoïde:

$$\phi = \pi r_0^2 H_0.$$

Ook dit merkwaardige resultaat roept de vraag op: zou deze $\boldsymbol{A}$ wel een fysische betekenis hebben?

Wij zijn hiermee begonnen aan wat het doel van dit artikel is: een excursie door enige vectorvelden, waarbij telkens de rotatie van het vectorveld het bindende element is. In het bijzonder zullen we dus situaties beschouwen waarin een werveldraad of wervelbuis gegeven is. Binnen zo'n buis geldt rot $\boldsymbol{P} \neq 0$, terwijl daarbuiten rot P $= 0$ is. Een aantal fysische verschijnselen wordt door dit bindende element in één verband gebracht. In een viertal gevallen zullen we nu de karakteristiek, het ontstaan en de dynamica van de wervels bekijken en daarbij op analogieën wijzen. Het is uiteraard niet de bedoeling overal tot strenge bewijsvoering of uitvoerige behandeling te komen.

Vloeistofwervels

De leer van de vloeistofstroming is al in de vorige eeuw tot ontwikkeling gekomen, in het bijzonder de eenvoudige theorie, die ook model gestaan heeft voor de ontwikkeling van de theorie der elektriciteit.

Het kenmerk bij uitstek van een vloeistof is zijn betrekkelijk geringe viscositeit of inwendige wrijving. Viscositeit veroorzaakt irreversibele overdracht van impuls van plaatsen waar de vloeistofsnelheid groot is naar plaatsen waar deze klein is, verondersteld dat het snelheidsveld een deformerende werking op een in het oog gevat vloeistofelementje heeft. Bij uniforme, starre rotatie van een vloeistofelement, met hoeksnelheid $\boldsymbol{\omega}$, geldt voor het snelheidsveld ter plaatse rot $\boldsymbol{v} = 2\boldsymbol{\omega}$, wat niet tot viskeuze effecten leidt. Waar impulsoverdracht plaatsvindt mag men echter een afremmende kracht, wrijving en dissipatie verwachten.

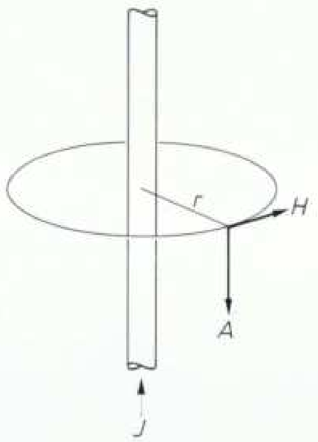


Fig. 1. Het magneetveld $\boldsymbol{H}$ en de vectorpotentiaal $\boldsymbol{A}$ in de ruimte rondom een oneindig lange draad waardoor een stroom $\boldsymbol{J}$ vloeit.

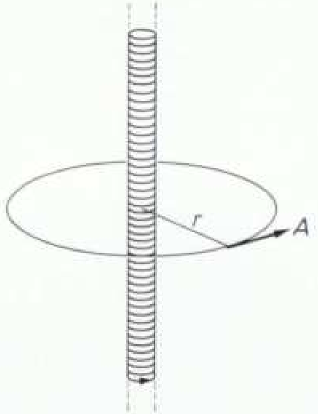


Fig. 2. De vectorpotentiaal $\boldsymbol{A}$ in de ruimte om een oneindig lange solenoïde.

In de ideale vloeistof kunnen we hiervan afzien. Als de gradiënt van de hydrostatische druk p de drijvende kracht is, geldt daar voor het vloeistofsnelheidsveld $\boldsymbol{v}$ de vergelijking van Euler (in navolging van de wet van Newton):

$$\varrho \frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = -\,\text{grad}\,p\,, \qquad (11)$$

waarin $\mathrm{d}\boldsymbol{v}/\mathrm{d}t$ de versnelling van een vloeistofelementje

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

is en ϱ de dichtheid ervan. Uit deze relatie kan de wet van Kelvin worden afgeleid:

$$\frac{\mathrm{d}}{\mathrm{d}t} \oint \boldsymbol{v}\, \mathrm{d}\boldsymbol{l} = 0 . \qquad (12)$$

Deze wet zegt dus dat de snelheidscirculatie $\oint \boldsymbol{v}\, \mathrm{d}\boldsymbol{l}$, genomen langs een gesloten, met de vloeistof meelopende contour, constant is. Als de vloeistof eenmaal een irrotationeel snelheidsveld heeft, blijft dit zo — althans in de benadering waarin de vergelijking van Euler geldt. Anderzijds blijft een eenmaal aanwezige wervel met onverminderde sterkte bestaan.

De karakteristiek van de vloeistofwervel is simpel. Als de rotatie van het snelheidsveld „zetelt" in een wervelas — en $\boldsymbol{v}$ dus overigens irrotationeel is — en er ook geen potentiaalstroom is, is het snelheidsveld concentrisch rondom die as, met

$$|v| = \frac{[c]}{r} , \qquad (13)$$

zie *fig. 3*. De analogie met fig. 1 is duidelijk. De situatie binnen in de wervel is minder simpel; in bepaalde gevallen kan er een holle ruimte zijn.

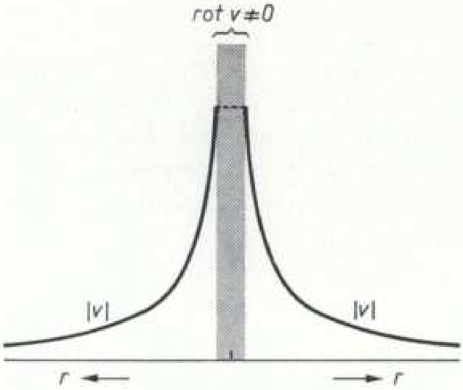


Fig. 3. Het snelheidsveld rondom een wervelas (grijze tint) in een ideale vloeistof. In de wervelas is rot $\boldsymbol{v}$ ongelijk nul.

Interessant is het geval dat de vloeistof niet isotroop is en het stel hydrodynamische parameters (zoals $\boldsymbol{v}$ en p) moet worden aangevuld met een zg. director $\boldsymbol{d}$. Dit heeft men in vloeibare kristallen; $\boldsymbol{d}$ kan dan bij voorbeeld de richting aangegeven waarin de lange moleculen, waaruit een vloeibaar kristal bestaat, lokaal zijn gericht. Geurst [2] heeft onlangs een uitbreiding van de stelling van Kelvin gegeven voor dit geval. Deze uitbreiding is het gevolg van het traagheidsmoment der moleculen voor een rotatie om een as loodrecht op $\boldsymbol{d}$. De boven behandelde circulatie $\Gamma = \oint \boldsymbol{v}\, \mathrm{d}\boldsymbol{l}$ heeft betrekking op de gewone zwaartepuntsbeweging van de moleculen. In de „directorruimte" is er echter ook een circulatie

$$\Gamma' = \alpha \oint \frac{\mathrm{d}\boldsymbol{d}}{\mathrm{d}t}\, \mathrm{d}\boldsymbol{d},$$

waarin α het zojuist genoemde traagheidsmoment is per eenheid van massa. Geurst heeft aangetoond dat $\Gamma + \Gamma'$ constant is.

[2] J. A. Geurst, Physics Letters **36A**, 63, 1971.

Hoe ontstaan wervels? Laten wij teruggaan naar het geval dat de vloeistof niet ideaal is. Nu geldt in plaats van de vergelijking van Euler (11), de vergelijking van Navier-Stokes, die in het geval van een incompressibele viskeuze vloeistof als volgt geschreven kan worden:

$$\varrho \frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = -\mathrm{grad}\, p - \eta\, \mathrm{rot\, rot}\, \boldsymbol{v} , \qquad (14)$$

waarin η de dynamische viscositeitscoëfficiënt is. De viscositeit veroorzaakt, bij voorbeeld, in een buis met uniforme cirkelvormige doorsnede bij een laminaire stroom het bekende parabolische snelheidsprofiel van Poiseuille:

$$v_x = \frac{r_0^2 - r^2}{4\eta} \frac{\mathrm{d}p}{\mathrm{d}x}, \qquad (15)$$

zoals onmiddellijk uit integratie van (14) volgt als we een stationaire toestand beschouwen.

Veronderstellen wij nu eens dat de viskeuze kracht op een vloeistofdeeltje (orde van de grootte $\eta v/r_0^2$) klein is vergeleken met de versnellingskracht op het deeltje als het zich eens van de wand loodrecht op de laminaire stroming zou begeven (orde van grootte $\varrho v^2/r_0$; hierbij is v de gemiddelde snelheid in de buis). Uit de analyse van fluctuaties in de vloeistofstroom blijkt dat, als deze versnellingskracht inderdaad verre overheerst boven de viskeuze, de laminaire Poiseuille stroming instabiel is. Er ontstaat turbulentie, met een min of meer duidelijke wervelstructuur. De beslissende grootheid is *Re*, het getal van Reynolds, dat is de verhouding van beide krachten:

$$Re = \frac{r_0 v}{\eta/\varrho}.$$

In andere situaties is er wel een andere afmeting l_{macr} kenmerkend. Altijd geldt:

$$Re = \frac{l_{\mathrm{macr}} v}{\eta/\varrho} , \qquad (16)$$

en de overgang van laminair naar turbulent vindt altijd plaats bij een waarde van *Re* die boven een zekere kritische waarde Re_{krit} ligt. Voor eenvoudige configuraties is Re_{krit} ongeveer 10^3.

Het is instructief de atomaire betekenis van η in te voeren. Men kan altijd schrijven:

$$\eta = l_{\mathrm{micr}} v_{\mathrm{th}} \varrho , \qquad (17)$$

waarin v_{th} de thermische snelheid der atomen is en l_{micr} een van de temperatuur afhangende grootheid die de dimensie van een lengte en de orde van grootte van de interatomaire afstand heeft. Nu geldt:

$$\frac{v_{\mathrm{krit}}}{v_{\mathrm{th}}} = [c] \frac{l_{\mathrm{micr}}}{l_{\mathrm{macr}}} , \qquad (18)$$

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

waarbij $[c]$ een numerieke constante is, ongeveer gelijk aan de kritische waarde Re_{krit} van het getal van Reynolds.

Relaties zoals (18) zullen we ook in het hierna volgende hoofdstuk ontmoeten. Er staat telkens in de noemer van het linkerlid een fatale belasting die de typische samenhang van het systeem kan verstoren — in dit geval een snelheid gelijk aan de thermische moleculaire snelheid. Bij een sterk gereduceerde waarde daarvan vormen zich echter reeds wervels. De reductiefactor is, afgezien van een numerieke constante, de verhouding van een microscopische lengte, welke iets zegt over de atomaire situatie, en een kenmerkende afmeting van het systeem, macroscopisch gezien.

We willen nu weer terugkeren naar de vergelijking van Euler en de gevolgen van wervelbewegingen in de ideale vloeistof bezien. Als het snelheidsveld niet uniform is, verandert $\mathbf{v}$ langs de stroombanen als volgt:

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}\cdot\mathbf{grad})\,\mathbf{v}\,. \qquad (19)$$

Met behulp van de identiteit

$$\operatorname{grad} \tfrac{1}{2}v^2 = \mathbf{v}\times\operatorname{rot}\mathbf{v} + (\mathbf{v}\cdot\mathbf{grad})\,\mathbf{v}\,, \qquad (20)$$

wordt de vergelijking van Euler:

$$\frac{\partial \mathbf{v}}{\partial t} + \operatorname{grad}\left(\tfrac{1}{2}v^2 + \frac{p}{\varrho}\right) = \mathbf{v}\times\operatorname{rot}\mathbf{v}\,. \qquad (21)$$

In een stationair stromingsveld is $\partial\mathbf{v}/\partial t = 0$, en aangezien het rechterlid een vector is die loodrecht op $\mathbf{v}$ staat (dus geen component heeft in de $\mathbf{v}$-richting), is de gradiëntterm een vector die in de stromingsrichting geen component heeft. Dit is de wet van Bernoulli: langs een stroomlijn genomen, is $\frac{1}{2}v^2 + p/\varrho$ constant.

Nu is het echter ook interessant te kijken naar een quasistationaire situatie waarin $\partial\mathbf{v}/\partial t \neq 0$ is en waarin in het bijzonder het snelheidsveld bewegende wervels bevat. Natuurlijk is, gemiddeld over een voldoend lange tijd, $\partial\mathbf{v}/\partial t$ toch wel te verwaarlozen. Met deze benadering geldt voor twee punten A en B in de vloeistof:

$$\left(\tfrac{1}{2}v^2 + \frac{p}{\varrho}\right)_B - \left(\tfrac{1}{2}v^2 + \frac{p}{\varrho}\right)_A = \int_A^B [\mathbf{v}\times\operatorname{rot}\mathbf{v}]\,\mathrm{d}\mathbf{l}\,. \qquad (22)$$

Stel dat A en B gekozen zijn in gebieden waar v te verwaarlozen is. We zien dat er nu een drukverschil tussen A en B kan bestaan als het rechterlid ongelijk nul is, d.w.z. de vloeistofbeweging moet per seconde een nettohoeveelheid rotatie (men spreekt wel van „vorticiteit") over de verbindingslijn AB voeren [3]; zie *fig. 4*.

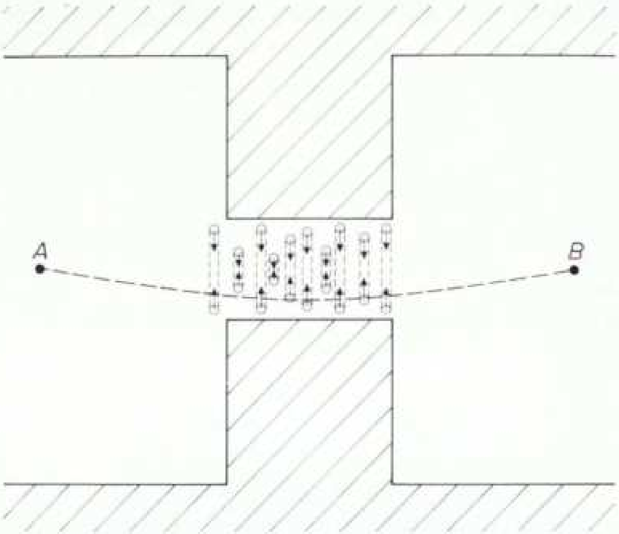


Fig. 4. Schematische voorstelling van de stroming van een vloeistof door een pijp met een sterk vernauwd gedeelte. In dit gedeelte ontstaan voortdurend wervelringen, die samenkrimpen (zie de pijltjes). Tussen A en B heerst een drukverschil dat correspondeert met de gestadige passage van wervels over de lijn AB.

De uitdrukking (22) is te vergelijken met soortgelijke vergelijkingen op het gebied van de elektromagnetische inductie; zie *fig. 5*. Tussen A en B bestaat in dit geval een potentiaalverschil dat gelijk is aan de gemiddeld per tijdseenheid over de lijn AB gevoerde magnetische flux. In het getekende voorbeeld [4] hebben we te maken met een unipolaire motor, maar in wezen gaat het zowel bij dynamo's als bij motoren altijd om de per seconde verzette magnetische flux.

Ten slotte kan men zich afvragen of de typische aspecten van de hydrodynamica ook kunnen worden teruggevonden in het geval van warmtestroming, in het bijzonder die in vaste stoffen. In het algemeen zal dit niet het geval zijn. Het voornaamste verschijnsel dat in dit verband de aandacht vraagt is een warmtestroom door fononen (pakketjes roostertrillingsenergie), die overwegend wordt geleid door botsingen van de fononen met de wand, dat is het oppervlak van het kristal [5]. Deze situatie is slechts bij zeer lage temperaturen en in zeer zuivere kristallen te realiseren. Het deeltjeskarakter van de fononen komt dan duidelijk tot uiting, en een Poiseuille-achtig snelheidsprofiel is mogelijk, maar van het ontstaan van warmtestroomwervels is niets bekend.

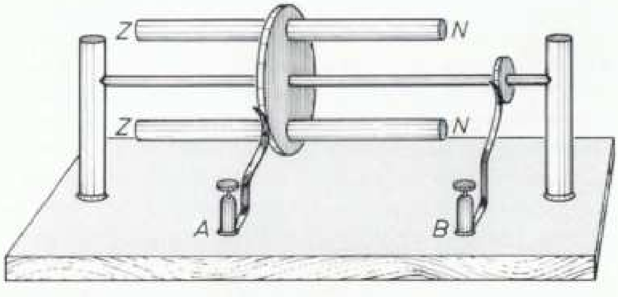


Fig. 5. Unipolaire motor. De aansluitklemmen A en B staan met elkaar in verbinding via de sleepcontacten en de draaibare koperschijf, waarin twee permanente magneten zijn gestoken. Een tussen A en B aangelegd potentiaalverschil vereist een voortdurend transport van magnetische flux — d.i. een rotationeel verdeelde stroom in de staven — over de verbindingslijn AB.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Dislocaties

Nadat we in het voorgaande op betrekkelijk klassiek gebied zijn gebleven, willen we nu een excursie maken in een gebied dat pas in deze eeuw tot ontwikkeling is gekomen, nl. dat van de dislocaties in kristallen. De vectorvelden die we beschouwen beschrijven nu geen stroming maar een verplaatsing of een vervorming.

Dislocaties veroorzaken diepgaande wijzigingen in de kristalorganisatie. Ten opzichte van het ideale kristal kan men het betrekkelijk eenvoudige geval van de *elastische* deformatie beschrijven door een kleine verplaatsing aan de roosterpunten te geven, zodanig dat deze verplaatsing overal een „nette" eenduidige functie van de plaatscoördinaten is. De verplaatsingsvector $\boldsymbol{t}$ is strikt genomen alleen in de roosterpunten gedefinieerd, maar men zou zelfs nog wel aan een continue functie $\boldsymbol{t}(x,y,z)$ kunnen denken, als men het gehele golfmechanische beeld van een uit kernen en elektronen bestaand kristal beschouwt. De symmetrie van het rooster is overal slechts zeer weinig verstoord; er is overal een één-op-één-afbeelding op het ideale rooster mogelijk. Maar in het ernstiger geval van *plastische* deformatie, gekenmerkt door een lijnvormige kristalfout die *dislocatie* wordt genoemd, is dit niet meer zo. Als nu een gesloten circuit wordt gekozen langs de roosterpunten *1*, *2*, *3*, . . . *N* in het afwijkende kristal en deze roosterpunten (die topologisch eenduidig ten opzichte van elkaar zijn gesitueerd) worden teruggezocht in het ideale kristal (*1′*, *2′*, *3′*, . . . *N′*), dan zal men soms merken dat het corresponderende circuit niet sluit en in dat geval omvat het circuit inderdaad een dislocatie (*fig. 6*). Om de kring te sluiten moet men een vector, gelijk aan een van de basisvectoren van het rooster, toevoegen ($N' \rightarrow 1'$). Hoe men de contour ook kiest, als het maar één dislocatie omvat is de ontbrekende stap altijd dezelfde. Dat is de zg. Burgers-vector, naar de hydrodynamicus J. M. Burgers, die een theoretische behandeling van dislocaties heeft gegeven lang voordat er een direct experimenteel bewijs was. De circulatie van de aldus geconstrueerde verplaatsingsvector heeft een eindige waarde.

Bij de beoordeling van de vervorming van een kristal interesseert ons eigenlijk niet hoe groot $\boldsymbol{t}$ is — een uniforme verplaatsing van het kristal als geheel leidt niet tot vervorming. Het gaat om de afgeleiden, die volgen uit een lokale vergelijking van de reële met de geïdealiseerde eenheidscel, en die de vervorming karakteriseren, zoals

$$\text{de dilatatie } e = \operatorname{div} \boldsymbol{t}$$

en

$$\text{de rotatie } \boldsymbol{\Omega} = \tfrac{1}{2} \operatorname{rot} \boldsymbol{t}. \qquad (23)$$

Deze moeten in een lineaire theorie volgen uit $\boldsymbol{f}$, de volumekracht (kracht per eenheid van volume), waarbij twee evenredigheidsconstanten een rol moeten spelen: de stijfheid tegen compressie en de stijfheid tegen afschuiven. De kracht $\boldsymbol{f}$ moet, zoals ieder vectorveld, af te leiden zijn uit twee potentialen:

$$\boldsymbol{f} = -\operatorname{grad} \phi + \operatorname{rot} \boldsymbol{\Psi}. \qquad (24)$$

Men krijgt dan:

$$e \sim \phi, \qquad \boldsymbol{\Omega} \sim \boldsymbol{\Psi}. \qquad (25)$$

De verplaatsing van de roosterpunten als gevolg van de invoering van een randdislocatie (fig. 6) is in het gehele kristal aanzienlijk. De verplaatsingsvector $\boldsymbol{t}$ heeft in dit geval overal vrijwel de richting van de Burgers-vector $\boldsymbol{b}$. Op grote afstand r van de dislocatie is t evenredig met de logaritme van b/r. De kristalsymmetrieën, die samenhangen met de anisotropieën in de bindingskrachten die het rooster bijeenhouden, maken de precieze berekening van de roostervervorming en de inwendige spanning ten gevolge van één enkele randdislocatie reeds tamelijk gecompliceerd. In het geval van een dicht netwerk van dislocaties is het onbegonnen werk. Het algemene beeld van de vervorming is dat de componenten ervan omgekeerd evenredig met de afstand tot een dislocatie afnemen. De situatie herinnert aan het magnetisch veld ($\sim 1/r$) en de genererende vectorpotentiaal ($\sim \ln 1/r$) rondom een lange stroomvoerende draad.

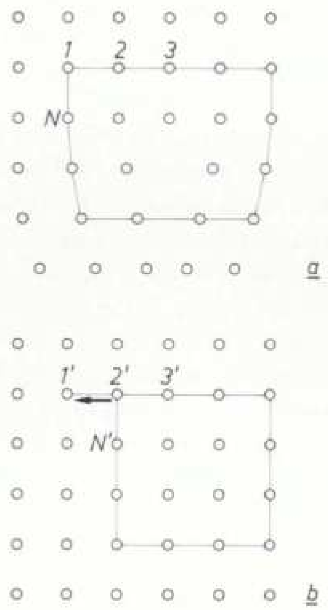


Fig. 6. Rondgang door het gestoorde rooster rondom een randdislocatie (*a*), afgebeeld op het ideale rooster (*b*). In het laatste geval vallen begin- en eindpunt niet samen. De extra stap (pijl) is de Burgers-vector.

[3] P. W. Anderson, Rev. mod. Phys. 38, 298, 1966.
[4] Gustav Wiedemann, Die Lehre von der Elektricität, 2e druk, 1893/98.
[5] H. B. G. Casimir, Physica 5, 495, 1938.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

We zullen nu het ontstaan en de beweging van dislocaties bespreken, waarbij we fysische verschijnselen zullen tegenkomen die hun pendant hebben in de fysica van de vloeistofwervels.

Dislocaties komen in beweging als plaatselijk een zo grote spanning ontstaat dat daar een atomaire reorganisatie plaatsvindt. Atomen worden doorgeschoven naar naburige eenheidscellen, en daarmee de dislocaties, maar het kristal breekt niet. Het proces van plastische deformatie van kristallen is eigenlijk alleen maar mogelijk dank zij de dislocatiebeweging!

Glazen hebben bij lage temperatuur een zeer grote viscositeit en missen beweeglijke dislocaties als hulpmiddel voor plastische vervorming, zodat glazen eerder breken dan vervormen. Ten aanzien van hun gedrag bij belasting zouden we de vaste stoffen kunnen onderscheiden in twee soorten: vaste stoffen van de eerste soort breken zonder een plastisch traject, vaste stoffen van de tweede soort kunnen plastisch meegeven dank zij de dislocaties en breken pas als het spel van de dislocatiebewegingen niet meer mogelijk is.

Dislocaties „vermeningvuldigen" zich volgens het Frank-Read-mechanisme uit initiële dislocaties, welke afkomstig zijn van groeifouten. De dislocaties vormen een netwerk, waarbij bepaalde dislocatiestukken tussen twee vaste punten kunnen uitbuigen en zo ten slotte nieuwe dislocaties afstoten [6]. Het is nu wel bekend hoe moeilijk het is dislocatievrije kristallen te kweken [7].

De afschuiving van een kristal zonder dislocatiebeweging zou een theoretische afschuifspanning τ_{th} vereisen van dezelfde orde van grootte als de afschuivingsmodulus. De geobserveerde kritische afschuifspanning τ_{kr} is vele ordes van grootte kleiner, dank zij de tussenkomst der dislocaties. Er geldt:

$$\frac{\tau_{kr}}{\tau_{th}} = [c]\,\frac{b}{l}, \qquad (26)$$

waarin b de Burgers-vector is en l de genoemde basislengte van de Frank-Read-bronnen, terwijl $[c]$ een dimensieloze constante is van de orde van grootte van 1. De verhouding b/l is de verhouding van een microscopische tot een macroscopische afstand, zoals we ook in (18) hebben ontmoet.

We zullen nu stilstaan bij het proces van de macroscopische vervorming, voor zover dit stationair verloopt. De afschuifsnelheid wordt dan bepaald door de concentratie σ van beweeglijke dislocaties met Burgers-vectoren b in de afschuifrichting, en hun snelheid v:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\Delta x}{x}\right) = bv\sigma. \qquad (27)$$

De reksnelheid van een staaf bij trekbelasting verkrijgt men uit (27) door vermenigvuldiging met een geometrische factor van de orde van grootte van 1, die rekenschap aflegt van de ligging der glijvlakken ten opzichte van de rekrichting [8].

Beschouwen we ten slotte een bijzonder geval — eigenlijk meer een gedachteproef. Stel dat afschuiving zou plaats vinden op een zeer ordelijke manier, dat wil zeggen op zodanige manier dat er slechts één dislocatie tegelijkertijd in het afschuifgebied — laat ons dit de „zwakke plek" noemen — was. Vermoedelijk zou, als het al lukken zou, een zeer geringe afmeting van dit gebied hiervoor gunstig zijn. Er is nu een frequentie f te onderkennen, nl. het aantal keren per seconde dat er een dislocatie passeert langs een gegeven punt. Het is de frequentie van een rikketik-effect dat gesuperponeerd is op de lineaire afschuiving en dat men met behulp van gevoelige mechanische of akoestische opnemers in principe zou moeten kunnen waarnemen. Het is wel waarschijnlijk dat een uitwendig aangekoppelde, afgestemde resonator hierbij zijn nut zou hebben. In formule:

$$\frac{\mathrm{d}}{\mathrm{d}t}(\Delta x) = fb. \qquad (28)$$

Omgekeerd is het denkbaar dat toepassing van een geluidstrilling een piek in de trek-reksnelheidskromme zou te zien geven bij juist die waarde van $\mathrm{d}(\Delta x)/\mathrm{d}t$ die door (28) wordt gegeven. We kunnen hier wijzen op de verregaande analogie met het Josephson-effect dat in het volgende hoofdstuk wordt behandeld.

De eigenschappen die kristallen ontlenen aan dislocaties zijn legio, maar het is niet de bedoeling hierover verder uit te weiden. We willen een uitzondering maken voor twee recente onderzoekingen. De eerste heeft betrekking op de reflectie van fononen aan het kristaloppervlak [9], een verschijnsel dat bij zeer lage temperaturen de warmtestroom in het kristal kan beheersen, zoals boven reeds vermeld werd [5]. De reflectie kan spiegelend of diffuus zijn. Gebleken is nu dat de diffuse variant ook kan ontstaan door een proces van multipele inelastische botsingen in een dunne laag vlak onder het oppervlak (enkele tientallen microns dik), wanneer zich daarin een flink aantal dislocaties bevindt. We zouden in dit geval eerder moeten spreken van absorptie en thermalisatie van de invallende fononen, gevolgd door reëmissie van fononen uit dit laagje.

Een ander onderzoek heeft aangetoond dat als gevolg van zo'n dislocatieconcentratie vlak onder het oppervlak ook een verbetering van de warmtestroom uit vloeibaar helium naar een vaste stof (b.v. een metaalfoelie in een warmtewisselaar) mogelijk is. We hebben hier een slechte aanpassing van de akoestische impedantie der beide media (de fononsnelheden en de dichtheden in de beide media verschillen respectievelijk een factor 20 en een factor 50). Daardoor is de reflectie-

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

coëfficiënt voor fononen groot en zal, wat nog erger is, een groot deel van de fononen totaal worden gereflecteerd. Maar doordat fononen die totaal zouden moeten reflecteren, uit het helium nog wel over een kleine afstand in het naburige medium binnendringen — de zg. evanescente golf —, kunnen ze daar verstrooid worden aan de dislocaties, met als gevolg een veel geringere reflectie. Dit is een interessante verklaring van de gevonden variaties in „warmteweerstand" (Kapitza-weerstand) van het oppervlak van verschillende koperfoelies. *Fig. 7* geeft recente berekeningen van dit effect door Haug en Weiss [10] weer.

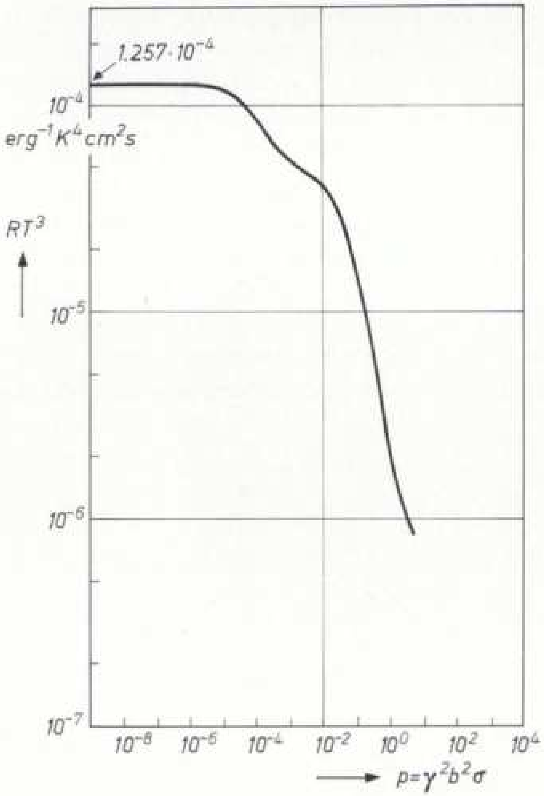


Fig. 7. De Kapitza-(warmte)weerstand R, zoals berekend [10] voor het geval dat evanescente roostergolven worden geabsorbeerd in een dun laagje dislocaties onder het oppervlak van een koperfoelie. Het produkt van R en de derde macht van de absolute temperatuur T is uitgezet tegen de grootheid p, die evenredig is met de concentratie σ der dislocaties (γ is een constante die ca. 2 bedraagt en b is de Burgers-vector).

Supergeleiders

Werveling van de stroom der ladingdragers in elektrische geleiders is, evenals de werveling der fononen in warmtegeleiders, een niet gemakkelijk te realiseren situatie. In geioniseerde gassen komt dit wel voor, tot op gigantische schaal zelfs, zoals bij voorbeeld uit foto's van het zonneoppervlak blijkt. Plasma-effecten in goed geleidende metalen leiden ook niet gemakkelijk naar het wervelconcept dat wij in dit overzicht als basis nemen. Deze wervels vinden we echter in optima forma in supergeleiders van de tweede soort. In het hiernavolgende vatten we even de voornaamste eigenschappen der supergeleiders samen [11].

In een supergeleider bevindt het gas van vrije elektronen zich in een zekere ordening. Door wisselwerking met het metaalrooster vormen deze elektronen paren, en wel op een zodanige wijze dat de totale gegeneraliseerde impuls,

$$\boldsymbol{p} = 2m\boldsymbol{v} + \frac{2e}{c}\boldsymbol{A}, \qquad (29)$$

van alle paren nul is — of dezelfde kleine waarde heeft ingeval er een transportstroom loopt. (In (29) is $\boldsymbol{v}$ de snelheid van het zwaartepunt van een paar.) In het golfmechanische beeld betekent dit dat de situatie beschreven wordt door een golffunctie Ψ die over macroscopische afstanden coherent is [*]. De situatie is te vergelijken met het elektromagnetische golfveld dat weliswaar uit ontelbare fotonen is samengesteld maar dat toch over macroscopische afstanden coherent kan zijn.

De golffunctie kan men schrijven als $\Psi = |\Psi| \exp i\phi$, waarbij $\boldsymbol{p}$ evenredig is met de golfvector in het Ψ-veld:

$$\boldsymbol{p} = \hbar \operatorname{grad} \phi .$$

Daar de rotatie van een gradiënt identiek nul is, is het $\boldsymbol{p}$-vectorveld in de supergeleider altijd irrotationeel:

$$\operatorname{rot} \boldsymbol{p} \equiv 0. \qquad (30)$$

Dit is het postulaat dat, in een iets andere vorm, reeds door de gebroeders F. en H. London ten grondslag is gelegd aan de elektrodynamische behandeling van de supergeleiders. Buiten het gebied van supergeleidende ordening kan rot $\boldsymbol{p} \neq 0$ zijn.

Vergelijking (30) leidt onmiddellijk tot de eerste vergelijking van London:

$$\operatorname{rot} \boldsymbol{j} = -\frac{ne^2}{mc}\boldsymbol{H}. \qquad (31)$$

Hierin is $\boldsymbol{j}$ de dichtheid van de stroom die gedragen wordt door de superelektronen en n hun concentratie; $\boldsymbol{H}$ is het lokale veld, dat uit de lokale vectorpotentiaal

[*] De vraag naar de maximale afstand waarover zich deze coherentie in een supergeleidende draad kan uitstrekken is behandeld door H. B. G. Casimir [12] en blijkt af te hangen van de diameter van de draad. Men zou het vermoeden kunnen uitspreken dat kristallografische coherentie in zeer dunne kristallijne draden over zeer grote afstanden óók niet mogelijk is.

[6] H. G. van Bueren, Philips techn. T. 15, 266 en 301, 1953, of H. G. van Bueren, Imperfections in crystals, North-Holland Publ. Co., Amsterdam 1961.
[7] B. Okkerse en P. Penning, Philips techn. T. 29, 116, 1968.
[8] J. J. Gilman, Mechanical behaviour of materials at elevated temperatures, McGraw-Hill, New York 1961, blz. 17.
[9] J. K. Wigmore, Physics Letters 37A, 293, 1971.
[10] H. Haug en K. Weiss, Proc. ICEC 4, Eindhoven 1972.
[11] J. Volger, Philips techn. T. 29, 1, 1968.
[12] H. B. G. Casimir, Proc. Kon. Ned. Akad. Wetensch. B 69, 223, 1966.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

$\boldsymbol{A}$ wordt afgeleid. Uit (31), gecombineerd met de Maxwell-vergelijkingen, volgt het Meissner-effect, dat is het verschijnsel dat H in de supergeleider nul is, behalve in een dun huidje onder het scheidingsvlak tussen de supergeleider en de aangrenzende niet-supergeleider. De indringing verloopt in eerste benadering exponentieel over een karakteristieke afstand λ, waarvoor geldt:

$$\lambda = \sqrt{\frac{mc^2}{4\pi ne^2}}. \qquad (32)$$

Als voorbeeld tekenen we in *fig. 8* de situatie binnen een supergeleidende staaf, waarin een niet te sterke stroom vloeit, die in het gehele volume dezelfde $\boldsymbol{p}$-waarde vastlegt. Merk op dat (31) een rotationele verdeling van de stroom in de indringdiepte tot uiting brengt. In de hieronder te bespreken situatie kan men echter pas echt van wervels of werveldraden in de supergeleiders spreken.

Neem aan dat binnen de supergeleider ergens een (niet-supergeleidend) gebied is waar rot $\boldsymbol{p} \neq 0$ is. Dat gebied moet zich dan uitstrekken langs een lijn of as die doorloopt tot aan het oppervlak of in zichzelf sluit, een zg. werveldraad. De waarde van $\boldsymbol{p}$ op een contour, rondom deze as gekozen, moet kennelijk ongelijk nul zijn. De interpretatie van $\boldsymbol{p}$ als golfvector brengt nu een zeer markant element in de behandeling: we zien dat als gevolg van de coherentie voor een dergelijke gesloten integratieweg rondom de bedoelde as geldt:

$$\oint \boldsymbol{p}\, \mathrm{d}l = nh, \qquad (33)$$

waarin h de constante van Planck is en n een geheel getal. Deze relatie was reeds bekend in de atoomtheorie van Bohr-Sommerfeld en komt nu terug bij de behandeling van het macroscopische quantumverschijnsel supergeleiding. In de niet-supergeleidende „boring" bevindt zich een zekere hoeveelheid magnetische flux. De bijbehorende kringstroom dringt slechts over de indringdiepte het omringende supergeleidende lichaam binnen; *fig. 9*. De flux is gequantiseerd, zoals uit (33) volgt:

$$\iint H \mathrm{d}S = \phi = n\phi_0, \qquad \phi_0 = \frac{hc}{2e} = 2\times 10^{-7}\,\text{gauss cm}^2.$$

De grootheid ϕ_0 is het zg. elementaire fluxquantum.

Het ontstaan van deze fluxdraden is een gecompliceerde zaak. We zien natuurlijk af van het triviale geval dat we in een supergeleider een fijn cilindervormig gaatje boren en hierin een hoeveelheid magnetische flux vangen. Waar het om gaat is dat in een grote klasse van supergeleiders, de zg. supergeleiders van de tweede soort, tijdens het magnetisatieproces fluxdraden spontaan worden gevormd aan het oppervlak en penetreren. Dit verschijnsel, dat wordt behandeld in de theorie van Ginzburg, Landau en Abrikosov, kan men duidelijk maken door terug te grijpen op de theorie van Gorter en Casimir [13]. In deze theorie wordt de thermodynamica van het magnetiseringsproces beschouwd en wordt voor het eerst ernst gemaakt met de supergeleidende toestand als fase met een eigen vrije energie F_{so}, die een bepaald bedrag kleiner is dan de vrije energie F_n van de normale fase. In het magnetisch veld komt bij F_{so} een additionele term $H^2/8\pi$ per volume-eenheid, wegens het perfecte diamagnetisme van de supergeleider, zodat er een kritisch veld H_c kan worden afgeleid uit de conditie

$$H_c^2/8\pi = F_n - F_{so}. \qquad (34)$$

Het in (34) voorgestelde energiebedrag moet men opbrengen om de gepaarde elektronen te scheiden.

Bij de beoordeling van het evenwicht tussen twee fasen moet men ook de oppervlakte-energie van de scheidingswand in aanmerking nemen. De onvolkomen verdringing van het magneetveld aan de buitenkant van de supergeleidende fase verlaagt de vrije energie. Een fijn verdeeld mengsel van normale en supergeleidende gebiedjes met een zo groot mogelijk contactvlak zal, als toestand van laagste vrije energie, inderdaad worden gerealiseerd, maar slechts bij die supergeleiders die aan het oppervlak een geringe verstoringsdiepte van de

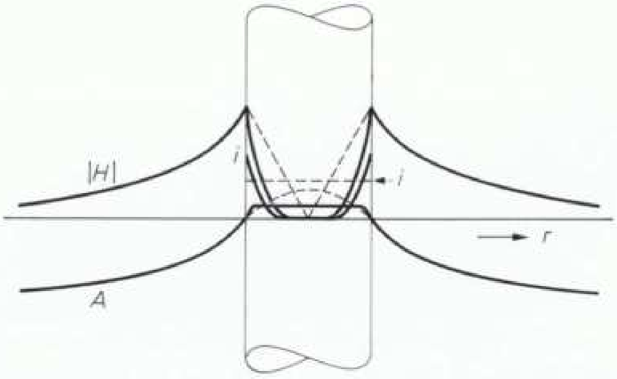


Fig. 8. De stroomdichtheid i, de vectorpotentiaal $\boldsymbol{A}$ en het magneetveld $\boldsymbol{H}$ bij een lange rechte supergeleidende draad. Buiten de draad hangt het ruimtelijk verloop van $\boldsymbol{H}$ en $\boldsymbol{A}$ er niet van af of de draad al of niet supergeleidend is, maar erbinnen wel. De gestippelde lijnen gelden voor de normale toestand. Op het oppervlak is $\boldsymbol{A}$ nul gesteld. De indringdiepte is overdreven groot getekend.

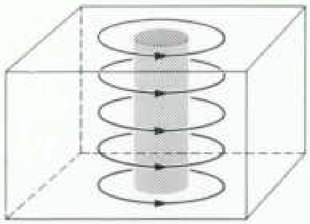


Fig. 9. Een wervel in een supergeleider. Deze bestaat uit een draadvormig normaal gebied (grijze tint) aan de wand waarvan een kringstroom loopt, die over een korte afstand in het omringende supergeleidende materiaal doordringt.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

supergeleidende ordening en daardoor een geringe positieve bijdrage tot de oppervlakte-energie vertonen. Dit is met name het geval bij legeringen, waarvoor in goede benadering geldt dat de verstoringsafstand ξ gelijk is aan de (geringe) vrije weglengte van de elektronen. De magnetische flux penetreert als een verzameling van zeer vele van de bovengenoemde wervels, met als bijzonderheid dat voor al deze werveldraden $n = 1$ is. De wervelas kunnen we opvatten als een cilindrisch gebiedje dat in de normale toestand verkeert en waarvan de straal de orde van grootte van ξ heeft. We spreken van de mengtoestand.

We zien hier, evenals in het geval van de dislocaties, dat de stijfheid van de ordening op een energetisch economische wijze wordt doorbroken door een invasie van wervels.

Wervelbeweging

Uit de thermodynamische analyse van de mengtoestand is gebleken dat er een kritische waarde H_{c1} van het magneetveld is, en dus ook een kritische waarde van de stroom aan het oppervlak, waarbij zich uit deze oppervlaktestromen werveldraden beginnen af te snoeren Het blijkt dat

$$\frac{H_{c1}}{H_c} = [c]\,\frac{\xi}{\lambda}, \tag{35}$$

waarin $[c]$ van de orde van grootte van 1 is. We kunnen ook kijken naar de kritische gemiddelde snelheid v_{cr} der elektronen in de dunne stroomvoerende laag. Het blijkt dat (35) herschreven kan worden als

$$\frac{v_{cr}}{\delta v} = [c]\,\frac{\xi}{\lambda}, \tag{36}$$

waarin δv de snelheidsvermeerdering is die men een elektron moet geven om de paarvorming ongedaan te maken. Opvallend is dat weer een intrinsieke grens die aan de „stijfheid" van de ordening is gesteld, gereduceerd wordt door het ontstaan van een wervelstructuur. De reductiefactor wordt gegeven door de verhouding van twee karakteristieke afstanden; in de noemer de „afmeting" van de „stroombuis", in de teller een vrije weglengte, vergelijk (18) en (26).

Bij supergeleiders is de magnetische plasticiteit, dat wil zeggen het tot stand komen van de mengtoestand, een zaak met twee kanten, een gunstige en een ongunstige. Dank zij de werveldraden is het mogelijk supergeleiding — als mengtoestand — tot bij opvallend hoge en dus technisch interessante waarden van de magnetische veldsterkte te behouden. Anderzijds echter hebben we te maken met de beweeglijkheid van de werveldraden, in het bijzonder wanneer er een transportstroom door de supergeleider wordt gezonden. Deze stroom oefent een kracht op de magnetische flux in de werveldraden uit waardoor deze gaan bewegen, en wel in eerste benadering loodrecht op de stroomrichting. Fig. 4 toont een situatie die ook in een draad van een supergeleider van de tweede soort kan voorkomen [14]. Voortdurend zullen er concentrische wervelringen aan het oppervlak van de draad gevormd worden en samensnoeren onder invloed van de stroom. Dit verschijnsel, dat overigens onder verschillende gedaanten voorkomt en flux flow wordt genoemd, leidt er toe dat langs de draad een spanningsdaling kan worden gemeten. In feite is er een weerstand, ondanks het feit dat we met een supergeleider te maken hebben. De grootte van de weerstand, d.w.z. de grootte van de spanning, kan worden berekend door het inductieverschijnsel te beschouwen dat wordt teweeggebracht door het bewegende inwendige fluxpatroon [15]. Een technisch interessantere variant van de supergeleiders van de tweede soort (de zg. harde supergeleider) is gekenmerkt door een grote onregelmatigheid in de kristalbouw (samenstellingsfluctuaties in legeringen, een fijn verdeelde tweede fase, roosterfouten van diverse soorten), zodat de fluxdraden hechtingsplaatsen kunnen vinden en zich niet meer bewegen, zelfs als er een sterke stroom door de draad gaat. We hebben hier te doen met een technologisch probleem, verwant met het harden van metalen en de „hardheid" van ferromagnetische materialen. Dit technologische probleem is goeddeels opgelost, zodat de toepassing van supergeleiders in de sterkstroomtechniek naderbij is gekomen. Grootscheepse toepassingen zullen echter gebonden zijn aan de betrouwbaarheid en het rendement van koelinrichtingen, d.w.z. aan de vooruitgang in de kryotechniek.

Keren we nog weer terug naar het verschijnsel van de flux flow. Het discrete karakter van de fluxbeweging stelt een statistisch probleem: in de weerstand van een supergeleider van de tweede soort zal ruis optreden, zoals door verscheidene onderzoekers ook is gevonden [16]. Een interessant geval is voorhanden als het gebiedje waar flux flow plaatsvindt van geringe afmeting is. Het blijkt dan soms mogelijk te zijn het totale aantal doorgangen van fluxdraden per seconde als een frequentie te bepalen. Voor deze frequentie f geldt:

$$V = f\,\frac{hc}{2e}, \tag{37}$$

waarin V de spanning over het gebiedje is, afhankelijk van de stroom. Met een geschikte detector kan men hf.

[13] C. J. Gorter en H. B. G. Casimir, Physica 1, 306, 1934.
[14] W. F. Druyvesteyn en J. Volger, Philips Res. Repts. 19, 359, 1964.
[15] A. G. van Vijfeijken, proefschrift, Amsterdam 1967 (ook gepubliceerd als Philips Res. Repts. Suppl. 1968, no. 8).
H. B. G. Casimir, Physics Letters 17, 177, 1965.
[16] G. J. van Gurp, proefschrift, Eindhoven 1969 (ook gepubliceerd als Philips Res. Repts. Suppl. 1969, no. 5).

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

signalen van de frequentie f volgens (37) waarnemen [17] (*fig. 10*).

Fraai is het zeer bijzondere, maar experimenteel wel degelijk te verwezenlijken geval dat de „zwakke plek" zo klein is dat er telkens slechts één fluxquantum tegelijk in aanwezig is. De stroom heeft nu een hoogfrequente modulatie waarvan de dominante frequentie

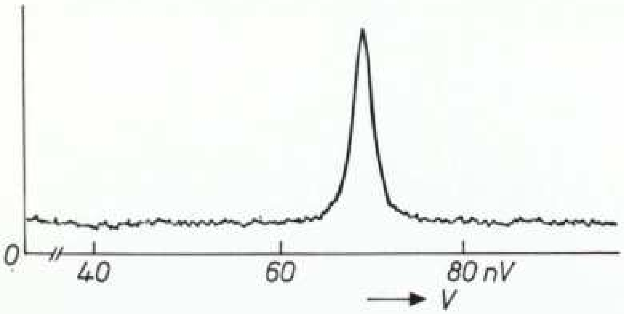


Fig. 10. Op 34 MHz gedetecteerd signaal, bij een spanning V van ca. 70 nV gegenereerd over een „zwakke plek" in een supergeleider van de tweede soort [17].

weer die van verg. (37) is. Dit is niets anders dan het beroemde dynamische Josephson-effect [**]. De zwakke plek kan bij voorbeeld zijn: een brugje of vernauwing in een supergeleidende metaalfilm, of een vastestofdiode met een zeer dun isolerend huidje als tussenmedium (ca. 30 Å), waardoorheen de supergeleidende golffunctie nog zijn coherentie kan bewaren. Er zijn experimenten gedaan waarbij de emissie van microgolfstraling is geconstateerd, maar ook is de invloed van geabsorbeerde elektromagnetische golven op de I,V-karakteristiek bestudeerd, waarbij is gebleken dat daarin pieken optreden bij die waarde van de spanning die door (37) wordt gegeven. Onlangs heeft Clark [18] de synchrone emissie (resp. absorptie) van microgolven waargenomen, afkomstig van een groot aantal Josephson-contacten die volgens een matrix waren gerangschikt.

Het experimenteren met de fragiele elementjes waarmee het Josephson-effect kan worden bestudeerd, heeft de laatste jaren een grote vlucht genomen. Tot slot van dit hoofdstuk vermelden we nog dat met behulp van het Josephson-effect zeer gevoelige stroommetingen kunnen worden gedaan in supergeleidende ketens. Het gaat in het kader van dit artikel te ver om dit uiteen te zetten, maar het verdient vermelding in verband met de vraag welke fysische betekenis de vectorpotentiaal heeft waarover we in het eerste hoofdstuk gesproken hebben. Beschouw een supergeleidende ring die een oneindig lange solenoïde omvat. Deze solenoïde kan een willekeurige, instelbare flux omvatten. Ter plaatse van de supergeleidende ring is geen magneetveld. Toch stelt zich, blijkens de metingen, in de ring een zodanig stroompje in dat aan de reeds door de solenoïde gegeven flux precies zoveel wordt toegevoegd dat de totale door de ring omvatte flux inderdaad een geheel aantal malen $hc/2e$ is. Het is alsof ter plaatse van de supergeleidende ring — dus daar waar geen magneetveld is — de vectorpotentiaal inderdaad wordt gevoeld. Een nauwkeurige behandeling van dit probleem vereist een beschouwing van alle energietermen waarmee het systeem supergeleidende ring + solenoïde + bekrachtiging wordt beschreven [19].

Vloeibaar helium II

Helium verkeert bij normale druk beneden de temperatuur van 2,1 K in een geordende toestand (He II), die in veel opzichten met de supergeleidende toestand van metalen kan worden vergeleken. Voor een deel zijn de atomen in een superfluïde fase gecondenseerd, met als merkwaardig gevolg dat de vloeistof althans voor geringe snelheden geen viscositeit heeft. Ook hier is de fysische situatie gekenmerkt door één golffunctie die zich coherent over het gehele vloeistofvolume uitstrekt. De ordening of stijfheid heeft voor het hydrodynamische gedrag van het superfluïdum de betekenis dat overal rot $\boldsymbol{v}_s = 0$ is, behalve op de as van de wervels die in het superfluïdum kunnen voorkomen. Weer geldt, evenals in het geval van de supergeleiders, de relatie van Bohr-Sommerfeld:

$$\oint \boldsymbol{p}_s\, \mathrm{d}\boldsymbol{l} = nh. \qquad (38)$$

Voor de wervels is $n = 1$ de normale situatie, zodat voor de circulatie K van het snelheidsveld $\boldsymbol{v}_s$ bij een wervel geldt:

$$K = \oint \boldsymbol{v}_s\, \mathrm{d}\boldsymbol{l} = \frac{h}{m}. \qquad (39)$$

Wervels ontstaan gemakkelijk in He II, en men zou het met recht een superfluïdum van de tweede soort mogen noemen. In een buis met een diameter van 1 mm is de kritische snelheid van de vloeistofstroom slechts ca. 0,1 cm/s. De gevormde wervels zijn vermoedelijk concentrische ringen, die aan het binnenoppervlak van de buis geboren worden. De hydrodynamica van He II is wegens de zo gecompliceerde verstrengeling met de statistische en golfmechanische aspecten van deze wonderlijke vloeistof nog niet volgroeid, maar enkele fenomenen zijn toch wel al tamelijk duidelijk.

De wervels vertegenwoordigen een hoeveelheid kinetische energie E. De berekening hiervan verloopt uiteraard geheel analoog met de berekening van de energie van het magnetische veld van een draad. De diameter $2a$ van de kern van een wervel in superfluïde helium is waarschijnlijk slechts enkele ångstroms groot.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Voor wervelringen waarvan de straal r_0 groot is ten opzichte van a, geldt met goede benadering:

$$E = \tfrac{1}{2}\varrho_s K^2 r_0 \left(\ln \frac{r_0}{a} + 0{,}33\right). \qquad (40)$$

Hierin is ϱ_s de dichtheid van het superfluïdum. Voor het magnetische geval geldt [20]:

$$E = \tfrac{1}{2} J^2 r_0 \left(\ln \frac{r_0}{a} + 0{,}33\right). \qquad (41)$$

(Als de vorticiteit (resp. de stroom) inhomogeen over de wervelas (de draad) verdeeld is, heeft de constante term een andere waarde dan 0,33.)

Wervelringen bewegen zich spontaan door de wisselwerking met de wand van het vat waarin de vloeistof zich bevindt [21]. Maar ook één wervelring die zich in een oneindig uitgestrekte vloeistof bevindt, verplaatst zich doordat een wervelring zich, om zo te zeggen, in zijn eigen snelheidsveld bevindt. Wat er precies gebeurt bij de wand als de wervelring zich vormt en een eigen leven gaat leiden, is nog een onderwerp van onderzoek [22]. Het is wel al gebleken dat voor de kritische snelheid v_{cr} in een buis met diameter d in eerste benadering geldt:

$$v_{cr} \approx \frac{K}{2\pi d}. \qquad (42)$$

Als we op het oppervlak van de wervelas voor v_s de intrinsieke maximale snelheid v_0 van de vloeistof aannemen, is (42) te schrijven als:

$$\frac{v_{cr}}{v_0} = [c]\frac{a}{d}, \qquad (43)$$

waarin $[c]$ van de orde van grootte van 1 is. Weer zien we de verhouding van een microscopische en een macroscopische lengte optreden in het rechterlid. Het experimentele onderzoek bevestigt verg. (42) vrij goed. Het verschijnsel van wervelvorming in He II is evenwel erg gecompliceerd, en in het bijzonder ten aanzien van wervelringen met zeer kleine diameter moet nog veel werk gedaan worden.

Op dit punt aangekomen zijnde zal het wel niemand verbazen dat in superfluïde helium verschijnselen zijn gevonden die analoog zijn met het dynamische Josephson-effect. Als er een drukverschil (en dus een hoogteverschil H) bestaat tussen twee compartimenten, gevuld met een superfluïdum en gescheiden door een wand met een klein gaatje („zwakke plek"), gaat dit gepaard met een regelmatige stroom van wervelringen langs dit gaatje. De overdwars per seconde getransporteerde vorticiteit is gelijk aan een frequentie f maal het circulatiequantum, zodat verg. (22) leidt tot:

$$mgH = hf, \qquad (44)$$

waarin g de versnelling van de zwaartekracht is. In *fig. 11* is een dergelijke experimentele opstelling geschetst [23].

De relatie (22) vereist niet alleen de zojuist besproken finesse ten aanzien van het quanteus karakter van de vorticiteit, maar ook is het nodig nog een generalisatie aan te brengen ten aanzien van het gebruik van de hydrostatische druk p. In het algemenere formalisme moet in plaats van p de thermodynamische potentiaal μ worden gebruikt; de drijvende kracht op de vloeistofdeeltjes is grad μ. De betekenis van deze generalisatie is van belang omdat we nu ook het effect van een tempera-

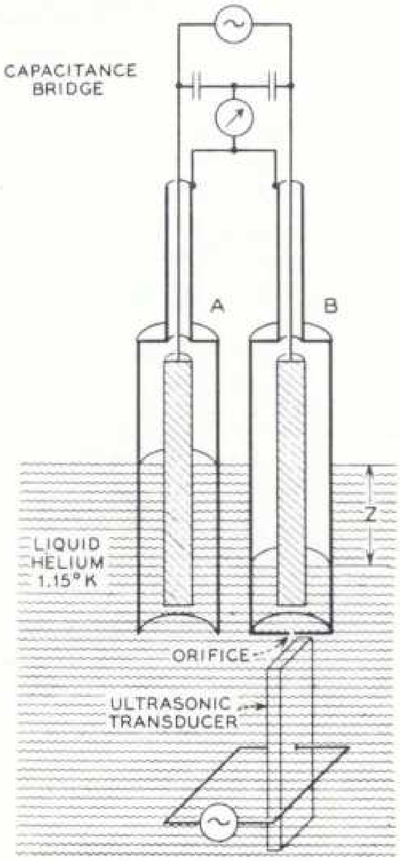


Fig. 11. Schets van een experiment met superfluïde helium, waarbij de vloeistofhoogte in de beide compartimenten capacitief wordt gemeten en de stroom door het gaatje wordt gemoduleerd door een generator van ultrageluid [23].

[**] Ten gevolge van niet-lineariteiten zijn eveneens harmonischen en subharmonischen aanwezig.

[17] R. K. Kirschman, H. A. Notarys en J. E. Mercereau, Physics Letters **34A**, 209, 1971.
[18] T. D. Clark, Physica **55** (Superconductivity, Proc. int. Conf. Stanford 1969), 432, 1971.
[19] H. B. G. Casimir, in: A. de Shalit, H. Freshback en L. van Hove (red.), Preludes in theoretical physics, North-Holland Publ. Co., Amsterdam 1966.
[20] Lord Rayleigh, Proc. Roy. Soc. A **86**, 562, 1912.
[21] A. Walraven, Phys. Rev. A **1**, 145, 1970.
[22] A. G. van Vijfeijken, A. Walraven en F. A. Staas, Physica **44**, 415, 1969.
[23] P. L. Richards en P. W. Anderson, Phys. Rev. Letters **14**, 540, 1965.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

tuurgradiënt in de beschouwing kunnen betrekken, en hiermee willen we besluiten.

Werveling kan in het geval van superfluïde helium worden onderdrukt door de vloeistof te brengen in een buis die gevuld is met een fijn-poreus materiaal. Men spreekt wel van een superlek. De thermodynamische potentiaal in een superlek heeft overal dezelfde waarde, omdat het rechterlid van (22) nul is, zodat geldt:

$$\mathrm{grad}\,\mu = \frac{1}{\varrho}\,\mathrm{grad}\,p - S\,\mathrm{grad}\,T = 0\,. \qquad (45)$$

Hierin is S de entropie per massa-eenheid. Bij een temperatuurverschil over het superlek stelt zich volgens (45) een drukverschil in, zoals London reeds had ingezien. Dit drukverschil kan tot een fonteineffect leiden. De krachten die de wervels in de poreuze structuur intomen worden wel vergeleken met de krachten die de fluxdraden in de harde supergeleiders binden aan de onregelmatigheden.

In goede benadering kan men stellen dat de stroom van superfluïde helium door een superlek slechts het in de grondtoestand gecondenseerde deel van de vloeistof verplaatst, hetgeen betekent dat er *geen entropietransport* is. Maar in een capillair waarin wervels ontstaan is grad $\mu \neq 0$, volgens (22). Hier zal ook het *niet* in de grondtoestand gecondenseerde, d.w.z. het energie- of entropiedragende, deel worden verplaatst. Als gevolg hiervan zullen in een serieschakeling van een superlek en een capillair waarin de kritische snelheid wordt overschreden, interessante thermische verschijnselen plaatsvinden. De situatie is te vergelijken met die in een Peltier-element, waar twee verschillende elektrische geleiders in serie zijn geschakeld, met andere woorden, waar een verschil is in gemiddelde energie van de stromende ladingdrager.

De vloeistof die uit het superlek stroomt kan men beschouwen als hebbende een absolute temperatuur van vrijwel nul graden Kelvin. Deze eigenschap zou men kunnen gebruiken. In de koelmethode van Mendelssohn [24] wordt deze van entropie gezuiverde vloeistof in een vaatje opgevangen; ze is inderdaad kouder (mechano-calorisch effect).

Wil men echter een continu werkend koeltoestel hebben, dan moet voortdurend warmte kunnen worden afgevoerd. De werveling in de afvoerstroom blijkt nu een

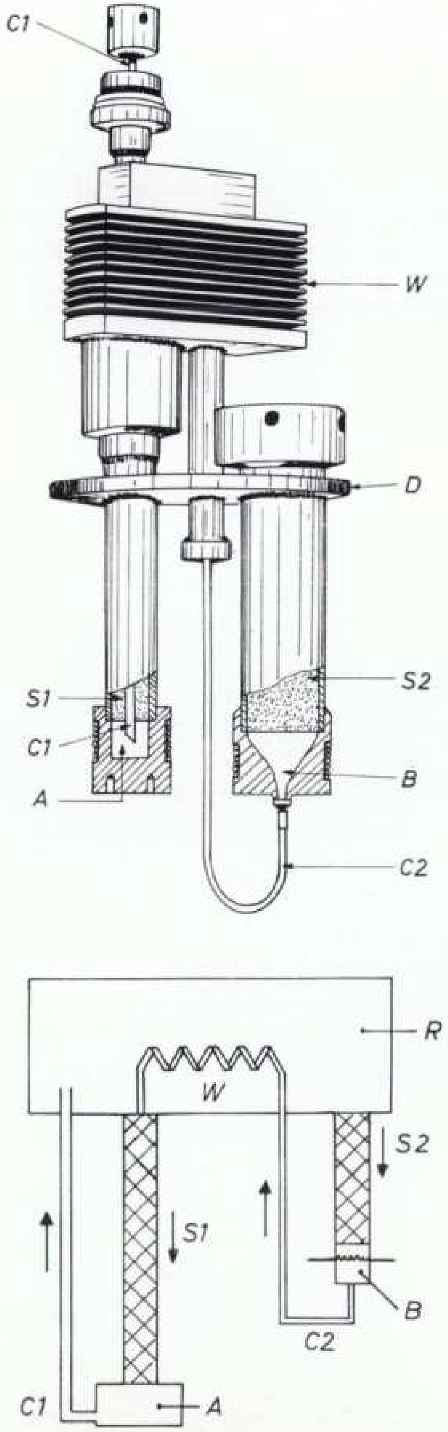


Fig. 12. Wervelkoeler volgens Staas en Severijns [25]. Het apparaat bevat twee ruimten A en B, die resp. via superlekken S_1 en S_2 en capillairen C_1 en C_2 met de warmtewisselaar W en met het heliumreservoir R van een kryostaat verbonden zijn. De wervelkoeler meet ca. 5×15 cm en is gemonteerd aan weerszijden van de bodem D van het heliumreservoir. De temperatuur van het vloeibare helium in R wordt door afpompen van de damp tot onder het λ-punt verlaagd (b.v. tot 1,5 K).

De ruimte B wordt elektrisch tot b.v. 1,7 K verwarmd, waardoor ten gevolge van het fonteineffect superfluïde helium vanuit R naar B stroomt. Het in B normaal geworden helium stroomt langzaam door capillair C_2 naar W. Daar wordt het weer superfluïde en kan dus via superlek S_1, ruimte A en capillair C_1 naar R stromen. De stroomsnelheid van het superfluïde helium in C_1 is zo groot dat er wervels ontstaan, waardoor er een drukgradiënt optreedt die het normale, entropie dragende helium uit A wegdrukt (d.w.z. er stroomt warmte van A naar R). Aangezien er alleen superfluïde helium (zonder entropie) naar A toestroomt, koelt de ruimte A af. Afhankelijk van de aan B toegevoerde warmte kan de temperatuur in A zeer nauwkeurig geregeld worden tussen 0,7 K en de temperatuur van het reservoir.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

gunstige omstandigheid, ja zelfs een noodzakelijke voorwaarde te zijn.

Een praktische uitvoeringsvorm is onlangs gepubliceerd door Staas en Severijns [25]. *Fig. 12* toont het essentiële deel. Zoals we zagen geldt in het superlek:

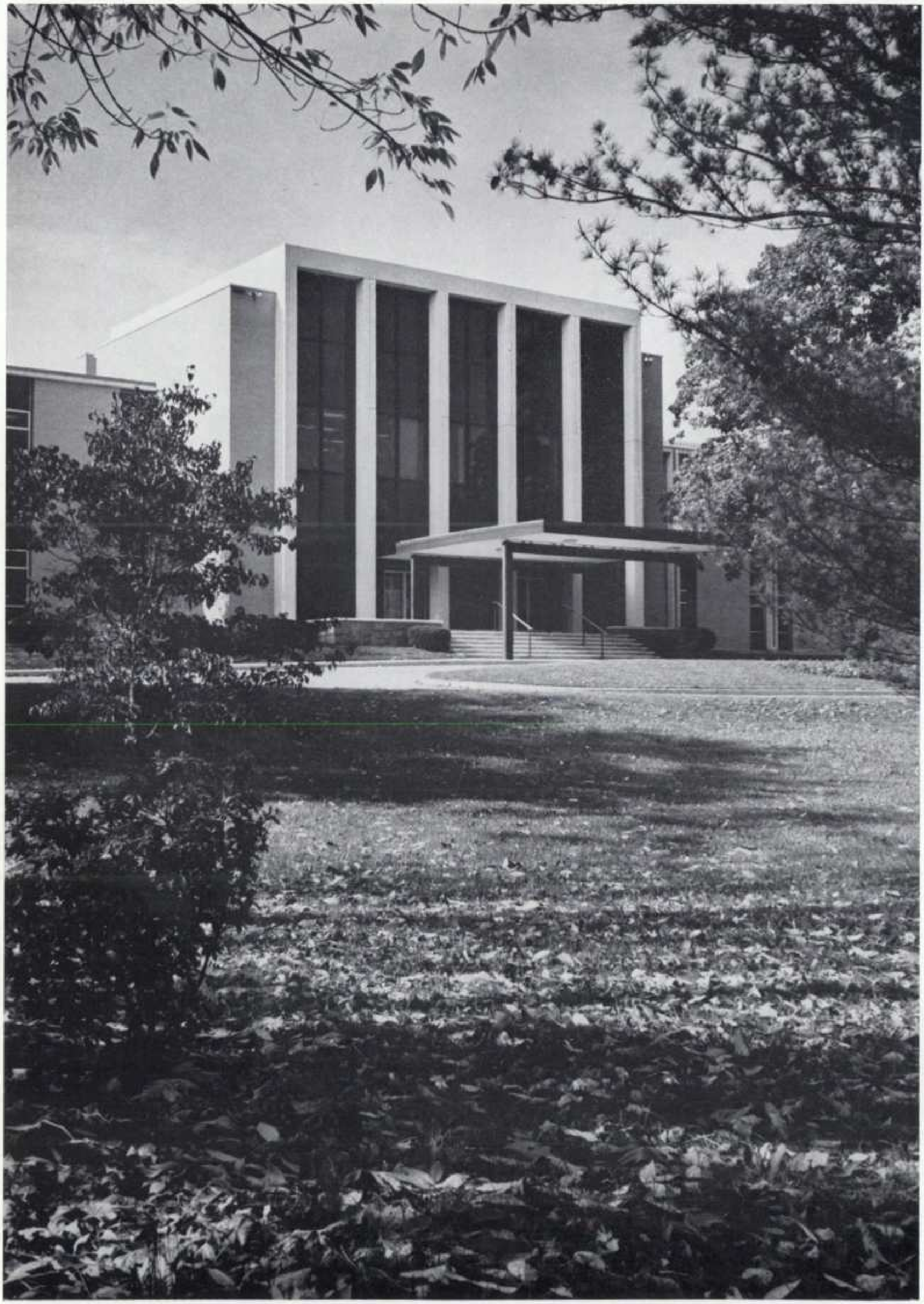
$$\Delta p = \varrho S \Delta T, \qquad (46)$$

d.w.z. dp en dT gaan gelijk op, maar in de afvoerbuis is grad $p \approx 0$ en stroomt de vloeistof tegen een temperatuurverval in. Het toestel wordt aangedreven met behulp van het fonteineffect — te vergelijken met het thermo-elektrisch effect (Seebeck-effect) tussen twee elektrische geleiders. Met behulp van dit koeltoestel, dat geen bewegende onderdelen heeft, kan men vooral in het temperatuurgebied tussen 2,0 en 0,7 K een heel behoorlijk koelvermogen produceren.

[24] J. G. Daunt en K. Mendelssohn, Nature **143**, 719, 1939.
[25] F. A. Staas en A. P. Severijns, Cryogenics **9**, 422, 1969, en Proc. ICEC 3, Berlijn 1970.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023



Philips Laboratories te Briarcliff Manor, N.Y., U.S.A.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

De fysica van licht emitterende centra in GaP

R. N. Bhargava

I. Inleiding

Een belangrijk nevenprodukt dat de technologie van vaste stof de laatste jaren heeft opgeleverd, is de vastestoflamp. De hoogste opbrengst aan zichtbaar licht verkrijgt men thans door deze lichtbronnen te maken van de III-V-halfgeleiderverbinding galliumfosfide. In dit artikel zullen we een overzicht geven van de fysica en de chemie van de luminescerende centra in GaP, waaraan deze lichtemissie met hoog rendement is te danken. De fysica van de luminescentie bestaat in het ontdekken en beschrijven van het mechanisme van de recombinatie van gaten en elektronen, die tot de emissie van licht leidt. De chemie houdt in: nagaan welke verontreinigingen in een kristal van belang zijn voor de fysische processen en vaststellen hoe zij toegevoegd of eventueel verwijderd moeten worden om luminescentie van een gewenste kleur te verwezenlijken. Alvorens we de fysische en chemische verschijnselen bij luminescerende centra in GaP gaan bespreken, zullen we in het kort een overzicht geven van diverse andere methodes voor lichtopwekking in vaste stoffen.

Licht kan in vaste stoffen op verscheidene manieren ontstaan. In gloeilampen, als eerste voorbeeld, wordt elektrische energie omgezet in straling door verhitting als tussenstap te gebruiken. De temperatuur en de grootte van de bron (b.v. een gloeispiraal van wolfraam) bepalen de lichtemissie van deze lampen. In lichtbronnen, gebaseerd op elektroluminescentie, kan de kinetische energie van elektronen evenwel rechtstreeks in licht worden omgezet zonder de extra complicatie van hoge temperaturen en vacuüm. De emissie-eigenschappen hangen af van de aard van het „moederkristal” en van de daaraan toegevoegde sporen van speciale verontreinigingen. Deze verontreinigingen kunnen centra doen ontstaan, die hetzij bruikbaar licht opwekken, hetzij nutteloos licht en warmte; deze centra zullen in het vervolg respectievelijk worden aangeduid als licht emitterende, of stralende, en niet-stralende centra.

Elektroluminescentie is voor het eerst waargenomen in 1923, door Lossew [1] [*]; in SiC-kristallen bleek in de buurt van een aangebrachte metaalelektrode licht te ontstaan. Later heeft Lossew kunnen vaststellen dat deze luminescentie afkomstig is van de recombinatie van (minderheids)ladingdragers, geïnjecteerd via ingebouwde *P-N*-overgangen. Een ander soort elektroluminescentie is waargenomen door Destriau [2], die de fosfor ZnS aan een wisselspanningsveld blootstelde door deze te suspenderen in een vloeistof waarin zich twee elektrodes bevonden. De optredende elektroluminescentie schijnt te ontstaan bij deëxcitatie van recombinatiecentra die aangeslagen zijn door niet-elastische botsing met thermisch vrijgemaakte ladingdragers; deze verkrijgen de vereiste kinetische energie door de versnelling die ze ondergaan in het elektrische veld. Men heeft aangetoond dat injectie van minderheidsladingdragers in halfgeleiders die tot luminescentie leidt, ook mogelijk is *zonder P-N*-overgang [3]. Alle op deze methodes gebaseerde opto-elektronische onderdelen en apparaten hebben in de praktijk tot dusver weinig succes gehad, òf omdat hun lichtopbrengst te laag is, òf omdat hun levensduur te kort is wanneer ze op het vereiste helderheidsniveau werken.

De ontdekking dat men met behulp van GaAs een vastestoflaser voor infrarode straling kan verwezenlijken [4] en dat men *P-N*-overgangen in GaP kan maken die met een zeer gunstig rendement rood licht emitteren [5] [6], heeft de belangstelling voor deze wijze van lichtopwekking sterk gestimuleerd. Sindsdien is er flinke vooruitgang geweest met betrekking tot lichtopbrengst en de mogelijkheid licht van verschillende kleuren op te wekken.

De op de *P-N*-overgang berustende vastestoflamp pleegt men aan te duiden met de afkorting LED (licht emitterende diode). Een LED bezit dezelfde voordelen ten opzichte van de gewone gloeilamp als de transistor tegenover de klassieke radiobuis; de LED's zijn uiterst betrouwbaar (bij levensduurexperimenten blijkt de halveringstijd $\gtrsim 10^6$ uur te zijn), ze bezitten een gunstig rendement (in het rode deel van het spectrum zelfs 12%) en ze werken op het lage vermogensniveau (milliwatts) dat goed past bij vastestofschakelingen. Zij worden dan ook veelvuldig toegepast in nieuwe apparatuur met vastestofelektronica, waarbij ze vele afbeeldingsfuncties vervullen, zoals die van verklikker-

Dr. R. N. Bhargava, medewerker van Philips Laboratories, Briarcliff Manor, N.Y., U.S.A.

[*] *De literatuuropgaven vindt men aan het eind van het artikel.*

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

lichtjes op cijfer- en letterpanelen, waarvoor tot nu toe neonlampjes enz. werden gebruikt.

In de laatste jaren is het verschijnsel van de luminescentie bij een *P-N*-overgang in een aantal overzichtsartikelen ter sprake gekomen [7-16]. In dit artikel beperken we ons daarom voornamelijk tot de licht emitterende en niet-stralende centra in GaP, omdat dit materiaal van wezenlijk belang is voor lichtbronnen die bij kamertemperatuur kunnen werken. De nadruk zal vallen op de recente ontwikkelingen betreffende emitterende centra waarvan men LED's met hoge lichtopbrengst kan maken.

gaan door absorptie binnen de halfgeleider, zodat dan geen licht naar buiten treedt. De beste definitie voor rendement is dan ook de verhouding van het aantal naar buiten tredende fotonen tot het aantal elektronen dat de stroom in de uitwendige keten draagt. Deze verhouding noemt men het uitwendige quantumrendement. Naast het in- en uitwendige quantumrendement bestaat ook een grootheid waarin het produkt voorkomt van het (uitwendige) quantumrendement en de gevoeligheid van het menselijk oog; men spreekt van lichtstroomrendement (Engels: luminous efficiency). Voor de toepassingen van een LED is waarschijnlijk

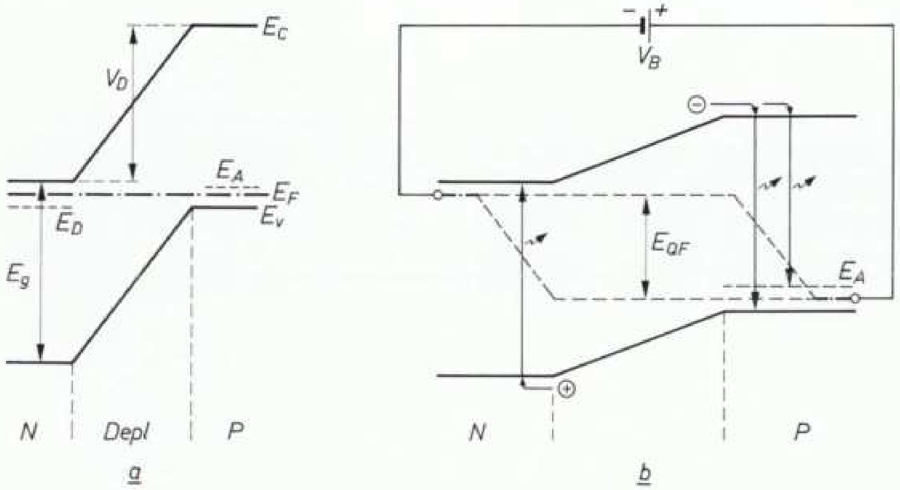


Fig. 1. Schematische voorstelling van een *P-N*-overgang in een halfgeleider, in de toestand van thermisch evenwicht, (*a*) zonder voorspanning, (*b*) met positieve voorspanning. V_D is de inwendige potentiaalsprong van de overgang; V_B is de aangelegde spanning; E_F is het Fermi-niveau; E_A en E_D zijn de niveaus van de acceptors en de donors; E_{QF} duidt de beide quasi-Fermi-niveaus aan, hun energieverschil is ongeveer gelijk aan eV_B. Elektronen, als minderheidsladingdragers, worden aan de *P*-kant van de overgang geïnjecteerd, gaten daarentegen aan de *N*-kant. Deze geïnjecteerde ladingdragers kunnen ter plaatse recombineren met de beschikbare gaten, respectievelijk elektronen, al of niet onder uitzending van straling. (Ontleend aan Dean [12].)

Elektroluminescentie kan bij een LED worden opgewekt door aan een *P-N*-overgang van de halfgeleider een spanning in de doorlaatrichting aan te leggen. In principe is het een eenvoudig proces, waarbij gaten afkomstig uit het *P*-gebied en elektronen uit het *N*-gebied in de grenslaagzone naar elkaar toe worden gedreven (zie *fig. 1*). Recombinatie van deze gaten en elektronen kan aanleiding geven tot de emissie van fotonen met een energie die gelijk is aan de bandafstand van de halfgeleider, of kleiner. Om zichtbaar licht te krijgen moet deze bandafstand groter dan 1,8 eV zijn. (Deze energie komt overeen met een golflengte van ongeveer 7000 Å.) Bij het recombineren wordt hetzij het gewenste licht opgewekt hetzij onzichtbare straling of warmte. Onder het inwendige quantumrendement verstaat men de verhouding van het aantal minderheidsladingdragers (gaten zowel als elektronen) dat bij recombinatie nuttig licht emitteert, tot het totale aantal dat recombineert. Een bij de recombinatie opgewekt foton kan verloren

dit lichtstroomrendement de belangrijkste grootheid. Door de ooggevoeligheid zullen een rood emitterende LED die een uitwendig quantumrendement van 3% heeft en een groen emitterende met een rendement van 0,1%, in een donkere kamer toch een even heldere indruk maken; het oog is immers voor groen licht ongeveer 30 maal zo gevoelig als voor langgolvig rood.

De luminescentie van kristallen bestudeert men in het algemeen met behulp van fotoluminescentie. Deze soort luminescentie van kristallen wordt opgewekt door absorptie van licht waarvan de fotonen een energie hebben die groter is dan de bandafstand. Voor een goede LED is een goed rendement bij zowel fotoluminescentie als bij injectie van ladingdragers nodig; in dit overzichtsartikel beperken we ons tot één van deze eisen, en behandelen we alleen de fotoluminescentie-eigenschappen van de halfgeleider. Volledigheidshalve merken we op dat er nog een excitatiemechanisme bestaat: bestraling van de halfgeleider met

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

elektronen, wat aanleiding geeft tot de zg. katodeluminescentie.

De tegenwoordig gebruikelijke *P-N*-overgangen voor emissie van infrarode en zichtbare straling maakt men van halfgeleiders als GaAs, GaP, GaAsP, GaAlAs, GaInP, en SiC. In de handel zijn LED's verkrijgbaar voor infrarood, rood, geel, en groen licht. Gedurende de laatste tien jaren is het inzicht in de fysica en de chemie van de luminescentie in deze halfgeleiders aanzienlijk toegenomen. De halfgeleiders GaAs en GaP hebben vooral veel aandacht gekregen vanwege hun hoge quantumrendement. De hoogste gerapporteerde waarde voor het uitwendige quantumrendement bij de GaAs-diodes (voor infrarood licht) is 32% [17]; voor GaP-diodes (rood licht) is dat 12% [18]. In wezen hangt het verschil tussen het quantumrendement bij GaAs en dat bij GaP op eenvoudige wijze samen met de waarschijnlijkheden van tot straling leidende overgangen zoals die bestaan bij halfgeleiders met een „directe" kleinste bandafstand en halfgeleiders met een „indirecte" kleinste bandafstand. We zullen dit punt bespreken in het volgende hoofdstuk.

Elektroluminescentie in GaAs ontleent zijn belang aan de ontwikkeling van een continu en bij kamertemperatuur werkende vastestoflaser [19]. Daarbij komt nog dat men van GaAs-diodes in combinatie met speciale fosforen, vastestoflampen voor verscheidene kleuren licht kan maken [20]. Een dergelijke fosfor, die gedoteerd is met zeldzame aardmetalen, kan de infrarode straling afkomstig van de GaAs-diode omzetten in zichtbaar licht, met andere woorden hij kan de frequentie opvoeren; in het algemeen zijn het fluorverbindingen, zoals YF_3 en YOF. Deze diodes met stralingsconversie zijn tegenwoordig om twee redenen aantrekkelijk: ten eerste zijn er al diodes gemaakt, zowel voor het blauwe als voor het groene als voor het rode gebied van het spectrum, en ten tweede verschaffen deze diodes de mogelijkheid de kleur binnen het zichtbare deel van het spectrum in zekere mate vrij te kiezen. Daar staat tegenover dat deze diodes een laag rendement bezitten. Het rendement van dergelijke fosforen met infraroodstimulatie is in de eerste plaats laag doordat de omzetting van het infrarood in zichtbaar licht gepaard gaat met de absorptie van twee of drie fotonen uit het infrarode deel van het spectrum, tegen een emissie van één foton van het zichtbare licht; zulk een proces heeft een lage overgangswaarschijnlijkheid. De tweede reden voor het lage rendement is de geringe absorptie van de door de GaAs-diode geëmitteerde infrarode straling in de desbetreffende fosfor (10% of minder). Om derhalve de fosfor aan een voldoend sterke lichtstroom bloot te stellen, moet men de GaAs-LED voeden met een bijzonder sterke stroom (100 mA of meer), mede doordat het rendement, naar gebleken is, sterk daalt wanneer de intensiteit van de exciterende straling wordt verminderd.

Anderzijds kan galliumfosfide rechtstreeks — en met een goed rendement — licht opwekken in het groene, het gele en het rode gebied van het spectrum. Het lichtstroomrendement van de diverse beschikbare diodes is vermeld in *tabel I*. De beste uitvoeringen zijn de GaP-diodes voor rood en groen, met een lichtstroomrende-

Tabel I. Het quantumrendement van een aantal diodes die zichtbaar licht uitstralen (LED's).

Materiaal		stralingsovergang: direct (d) indirect (i)	max.		quantumrendement	lit.	lichtstroomrendement (lumen/watt)	
			$h\nu$ (eV)	λ (nm)			bij quantumrendement 100%	beste gemeten waarde
$GaAs_{0,6}P_{0,4}$	(rood)	d	1,91	649	2×10^{-3}	I	75	0,15
$GaAs_{0,18}P_{0,82}$	(geel)	i	2,10	590	3×10^{-4}	II	≈ 500	≈ 0,15
$Ga_{0,65}Al_{0,35}As$	(rood)	d	1,84	675	$1,3\times10^{-2}$	III	≈ 35	≈ 0,45
$In_{0,4}Ga_{0,6}P$	(groen)	d	2,17	571	2×10^{-4}	IV	610	0,12
GaP	(rood)	i	1,77	699	12×10^{-2}	V	20	2,4
GaP	(groen)	i	2,22	558	6×10^{-3}	VI	600	3,6
GaP	(geel)	i	2,05	605	1×10^{-3}	VII	≈ 400	0,4
SiC	(geel)	i	2,1	590	3×10^{-5}	VIII	330	0,01
GaN	(groen)	d	2,4	515	1×10^{-2}	IX	100	0,01

I. A. H. Herzog, W. O. Groves en M. G. Craford, J. appl. Phys. **40**, 1830, 1969.
II. Lit. [68].
III. J. M. Blum en K. K. Shih, Proc. IEEE **59**, 1498, 1971.
IV. B. W. Hakki, J. Electrochem. Soc. **118**, 1469, 1971, en persoonlijke mededeling.
V. Lit. [18].
VI. Lit. [27].
VII. Lit. [65] en [52].
VIII. R. M. Potter, J. M. Blank en A. Addamiano, J. appl. Phys. **40**, 2253, 1969.
IX. Lit. [77]; men merke op dat in dit geval een vermogensrendement van 10^{-4} gebruikt is om het lichtstroomrendement te kunnen berekenen.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

ment van respectievelijk 2,4 lm/W en 3,6 lm/W. De rest van dit artikel is gewijd aan de wijze waarop deze luminescentie met hoog rendement in GaP is bereikt. Ten slotte zullen we wat zeggen over het in de toekomst wellicht bereikbare quantumrendement.

II. Bandstructuur

De kristalgroei en de elektrische eigenschappen, alsmede de voldoend grote bandafstand (2,26 eV bij 300 K), maken GaP geschikt voor het opwekken van zichtbaar licht. Galliumfosfide is echter een halfgeleider met een „indirecte" kleinste bandafstand. (In de zg. $\boldsymbol{k}$-ruimte liggen bij een halfgeleider met „directe" kleinste bandafstand, zoals GaAs, de minimale energie van de geleidingsband en de maximale energie van de valentieband bij dezelfde waarde van de golfvector $\boldsymbol{k}$, dus op één verticale lijn in het E,k-diagram. In het geval van een halfgeleider als GaP liggen deze extrema daarentegen bij verschillende waarden van k.) De bandstructuur van GaP, en ook die van GaAs, is afgebeeld in *fig. 2*. In een halfgeleider met niet-verticale kleinste bandafstand moet er bij de recombinatie van een elektron uit de geleidingsband met een gat uit de valentieband nog de hoeveelheid impuls in rekening gebracht worden die overeenkomt met $\boldsymbol{k} = \boldsymbol{k}_c - \boldsymbol{k}_v$, omdat voor de impuls een behoudswet geldt; de golfvectoren $\boldsymbol{k}_c$ en $\boldsymbol{k}_v$ hebben respectievelijk betrekking op het elektron in de geleidingsband en het gat in de valentieband. De intrinsieke recombinatiewaarschijnlijkheid voor de elektronen en gaten is klein ten gevolge van de beperking, opgelegd door het behoud van de kristalimpuls. Het recombineren van elektronen en gaten kan evenwel zeer bevorderd worden door de toevoeging van bepaalde verontreinigingen die een sterke wisselwerking vertonen met vrije ladingdragers; de golfvector $\boldsymbol{k}$ blijft dan behouden door de wisselwerking tussen ladingdrager en verontreiniging. Men kan zich dus voorstellen dat de genoemde verontreiniging een verbetering van het luminescentierendement teweeg brengt. Een dergelijk luminescentieproces via centra is verantwoordelijk voor de sterke lichtproduktie in GaP.

Enige eenvoudige voorbeelden van de vele mogelijke, door verontreinigingen geïnduceerde overgangen in GaP zijn: (a) recombinatie van een elektron dat gevangen is in een donorniveau, met een gat dat gevangen is in een acceptorniveau, (b) recombinatie van een door een „diep" donorniveau gevangen elektron met een vrij gat, (c) recombinatie van een exciton dat gebonden is aan een „iso-elektronisch" of een ander neutraal centrum. Een exciton is de combinatie van een elektron en een gat die als paar gebonden zijn. Bij een iso-elektronisch centrum heeft men te maken met een lokale gebonden toestand, ontstaan door een iso-elektronische substitutie (bij voorbeeld N, in plaats van P, in GaP). Een dergelijk centrum, dat elektrisch gesproken neutraal is, kan òf een puntfout zijn òf een molecuul in het GaP-rooster. Wanneer het iso-elektronische centrum een exciton bindt, kan dat leiden tot een recombinatie waarbij met goed rendement licht geëmitteerd wordt. Het beste voorbeeld van een roosterpuntfout is het geval van een iso-elektronisch centrum waarbij het fosforatoom in GaP door een stikstofatoom is vervangen (de zg. groene diode). Een welbekend iso-elektronisch centrum van het molecuultype is het naaste-buurcomplex Zn_{Ga}^{-}-O_P^{+}, dat verantwoordelijk is voor de rode emissie van GaP. Deze twee soorten iso-elektronische centra zullen respectievelijk in hoofdstuk IV en hoofdstuk V gedetailleerd besproken worden; bij kamertemperatuur zijn deze centra in GaP de voornaamste die door verontreinigingen geïnduceerde

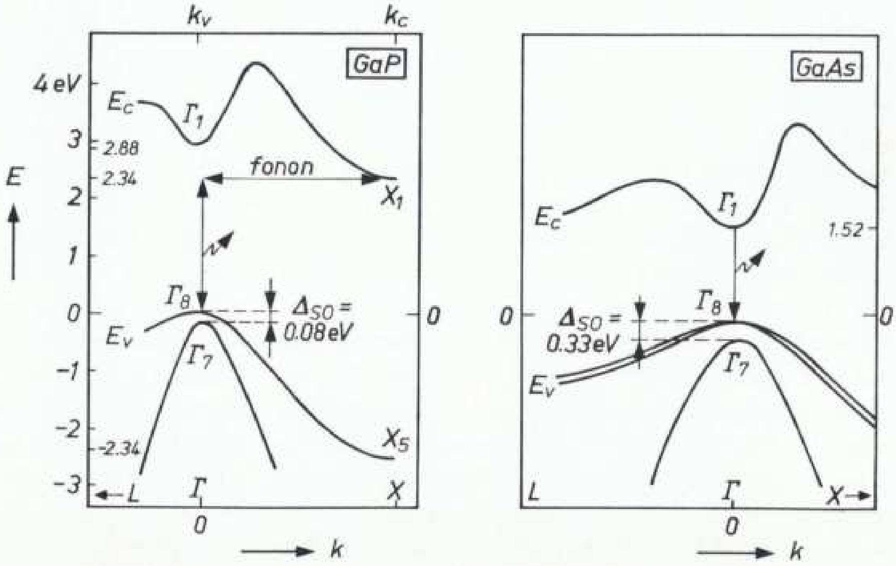


Fig. 2. *Links:* de bandstructuur van de halfgeleider GaP (een III-V-verbinding), met een „indirecte" kleinste bandafstand $X_1 - \Gamma_8$. *Rechts:* idem voor GaAs, met een „directe" kleinste bandafstand $\Gamma_1 - \Gamma_8$. In het geval van GaAs kunnen elektronen uit de geleidingsband rechtstreeks recombineren met gaten uit de valentieband, waarbij een foton vrijkomt. „Rechtstreeks" betekent dat de impuls zonder meer behouden blijft. Bij GaP geschiedt deze recombinatie met „fononassistentie", opdat de kristalimpuls behouden blijft. Het energieverschil Δ_{SO} bij de valentieband is te wijten aan de zg. spin-baan-koppeling.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

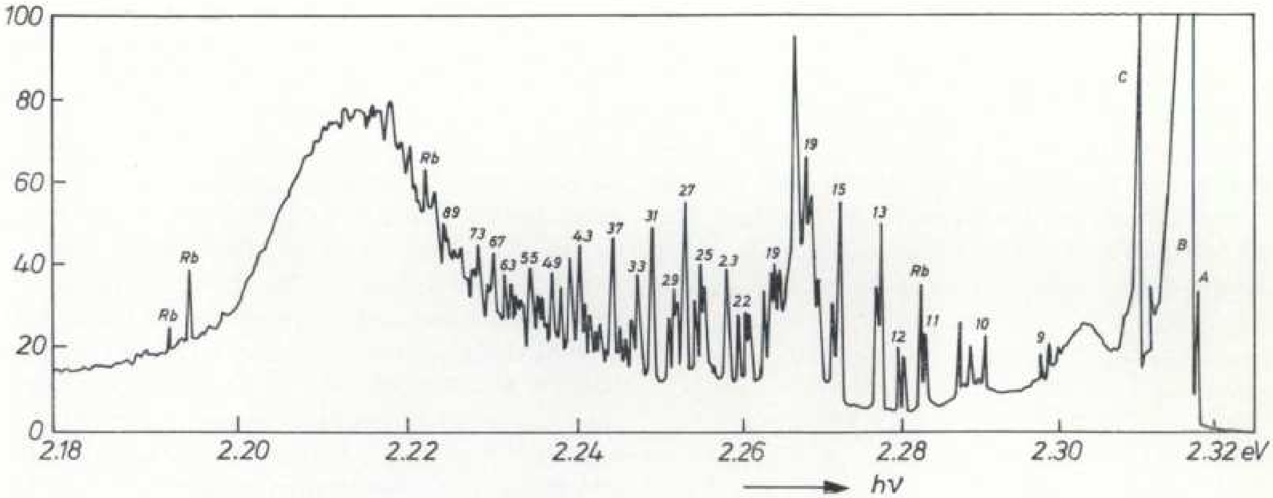


Fig. 3. Fotoluminescentiespectrum (intensiteit tegen de energie $h\nu$ van de fotonen) van GaP van nominaal zuivere kwaliteit, gemeten bij 1,6 K. Het spectrum is een zg. donor-acceptor-paar-spectrum (zie ook fig. 4). Een elektron, gebonden aan een zwaveldonor, recombineert onder het uitzenden van straling met een gat dat aan een koolstofacceptor gebonden is. De discontinuïteit in het verloop van de afstanden tussen de betrokken donors en acceptors leidt tot de lijnenstructuur. De getallen bij de lijnen slaan op de zg. schilnummers („lattice shell numbers"); de werkelijke afstanden hangen er op eenvoudige wijze mee samen. De lijnen *Rb* zijn ijklijnen van een rubidiumlamp. De lijn *C* ontstaat door de recombinatie van een exciton, gebonden aan een neutrale zwaveldonor; de lijnen *A* en *B* zijn afkomstig van een recombinatie van een exciton dat gebonden is aan het iso-elektronische centrum, gevormd door een stikstofatoom op een fosforplaats. (Ontleend aan Thomas [8].)

luminescentie geven. In hoofdstuk VI zullen we de problemen bespreken die samenhangen met de vorming van de neutrale naaste-buurcomplexen.

Het is bijzonder leerzaam niet alleen deze stoorcentra te bestuderen, waarbij de recombinatie met de emissie van gewenste straling samengaat, maar ook de concurrerende recombinatieprocessen die stralingsloos verlopen. Deze processen dissiperen energie, afkomstig van de geïnjecteerde elektronen, en verminderen dientengevolge de lichtopbrengst. Als voorbeeld het volgende geval: bij een temperatuur van 4 K komt in GaP het rendement van het recombinatieproces (a), de donor-acceptor-overgang, dicht in de buurt van 100 %. Voor elk geïnjecteerd elektron krijgen we dan gemiddeld ongeveer één foton. Bij kamertemperatuur (300 K), daarentegen, is het quantumrendement minder dan 0,01 %. Deze drastische daling van het rendement hangt samen met het feit dat de gevangen elektronen en gaten bij deze hogere temperatuur een aanzienlijke kans hebben langs thermische weg te worden losgemaakt van de acceptors en de donors, waarbij dan nog als extra factor de betrekkelijk geringe overgangswaarschijnlijkheid van D-A-paren in het spel treedt [78]. Zodra ze terug zijn in de geleidingsband, is het mogelijk dat stralingsloze recombinatie via andere processen optreedt.

Ook het grootste deel van de energie die in proces (c) vrijkomt, dus bij de recombinatie van een exciton gebonden aan een neutrale donor of acceptor, gaat verloren aan stralingsloze processen. De excitonen zijn echter zeer zwak gebonden aan deze centra, zodat we niet verwachten dat het nagenoeg stralingsloze proces (c) bij de hogere temperaturen enig gewicht in de schaal legt; daar winnen thermische effecten het. We zullen de verdere discussie van deze niet-stralende centra, die van veel belang zijn, uitstellen tot hoofdstuk VIII.

III. Donor-acceptor-paaremissie

Zoals reeds vermeld is, kan het recombinatiemechanisme waarbij de niveaus van eenvoudige, b.v. atomaire, donors en acceptors betrokken zijn, bij lage temperaturen met zeer goed rendement werken; maar niet bij kamertemperatuur. Toch heeft de bestudering van de tot emissie van licht leidende recombinatie bij deze niveaus een grote hoeveelheid kennis omtrent de aard van de donor- en acceptorcentra opgeleverd. We gaan deze recombinatieprocessen meer in details bespreken.

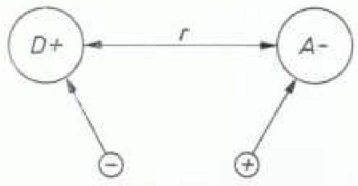
III-A. Paaremissie van ondiepe donor- en acceptorniveaus

Fig. 3 toont het fotoluminescentiespectrum van GaP gedoteerd met S en C, waargenomen bij een temperatuur van 1,6 K. De emissie vindt plaats in het groen. Het spectrum bestaat uit een groot aantal scherpe lijnen, die aan de kant van de lage energie samenvloeien tot een hoge en brede piek, een soort band. Hopfield c.s. [21] geven een eenvoudig model waarmee het bestaan van al deze lijnen verklaard kan worden. Zoals afgebeeld is in *fig. 4*, kunnen vrije elektronen ge-

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

vangen worden door positief geladen (geïoniseerde) donorcentra. Op overeenkomstige wijze kunnen negatief geladen acceptorcentra vrije gaten vangen. En, omdat de temperatuur laag is, zullen de gevangen elektronen en gaten blijven zitten waar ze zijn; er is niet genoeg thermische energie om ze door ionisatie naar geleidings- en valentieband terug te brengen. Dat op zijn plaats blijven is echter maar tot op zekere hoogte waar; elektronen en gaten hebben immers



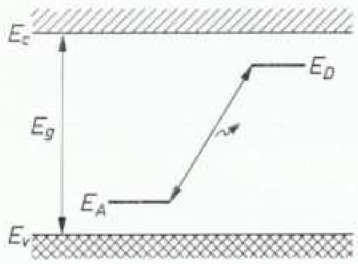


Fig. 4. *Boven:* schematische voorstelling van de vangst van een paar, bestaande uit een vrij elektron en een vrij gat, door een geïoniseerd donoratoom en een geïoniseerd acceptoratoom. *Beneden:* de daarop volgende recombinatie van de gevangen deeltjes leidt tot de zg. donor-acceptor-paaremissie. De energie $h\nu$ van de uitgezonden quanta worden kleiner als r toeneemt; de oorzaak van dit verschijnsel is de Coulomb-wisselwerking. (Ontleend aan Thomas [8].)

golffuncties die zich in het rooster uitstrekken tot buiten de verontreinigingsatomen. Hoewel de donors en acceptors zich op enige afstand van elkaar bevinden, zullen toch de elkaar overlappende golffuncties van een dergelijk elektron-gatpaar leiden tot een van nul verschillende waarschijnlijkheid voor recombinatie waarbij een foton zichtbaar licht ontstaat (zie fig. 4). De geëmitteerde hoeveelheid energie $E(r)$ is gelijk aan de bandafstand, verminderd met de bindingsenergieën van het acceptor- en het donorcentrum maar vermeerderd met een Coulomb-bijdrage ($e^2/\varepsilon r$, in zijn eenvoudigste vorm):

$$E(r) = E_g - (E_A + E_D) + e^2/\varepsilon r. \qquad (1)$$

In deze formule is E_g de bandafstand; E_A en E_D zijn de bindingsenergieën van respectievelijk acceptor-gat en donor-elektron; ε is de diëlektrische constante en r is de afstand in het kristal tussen het acceptor- en het donorcentrum. De Coulomb-bijdrage mag men opvatten als een correctie op de energie waarmee het elektron is gebonden aan het donoratoom; het elektron wordt immers niet naar het oneindige verplaatst, maar naar een acceptor-centrum op afstand r.

Een uitermate belangrijk aspect van bovengenoemde vergelijking is het feit dat het waargenomen spectrum uit vele scherpe lijnen bestaat; zij corresponderen met de verschillende discrete waarden voor r, die optreden doordat de acceptor- en de donoratomen discrete roosterplaatsen bezetten. Elke lijn uit het spectrum hoort dus bij zijn „eigen" paren atomen (donors en acceptors), die een voor de lijn karakteristieke afstand bezitten. Men verwacht dan dat de intensiteit van dergelijke „paarlijnen" evenredig is met het aantal paren dat de desbetreffende karakteristieke afstand heeft. De intensiteitsverdeling in het spectrum leert dat het aantal mogelijke paren niet regelmatig toeneemt bij stijgende karakteristieke afstand. Veronderstellen we nu dat de donor- en acceptoratomen willekeurig verdeeld zijn, dan kunnen we de intensiteitsverdeling van de paarlijnen voorspellen uit de bezetting van de betrokken boloppervlakken in het rooster, waarbij we natuurlijk eerst de precieze waarde van de schilstraal r voor elke lijn van het spectrum moeten hebben vastgesteld. (De groep naaste-buuratomen die een als centrum gekozen atoom omringen, vormen de „schil" (Engels: shell) in het kristalrooster met nummer 1, de daarna volgende buren de schil met nummer 2, etc.)

Zodra de waarden van r bekend zijn, kunnen we de optredende lijnenergie uitzetten als functie van de afstand r die de twee betrokken verontreinigingen kan scheiden. Extrapoleren van het gevondene naar $r = \infty$, verschaft ons dan met behulp van formule (1) de waarde van $E_g - (E_A + E_D)$. De bandafstand E_g is bekend, zodat één van de beide bindingsenergieën op deze wijze nauwkeurig bepaald kan worden als we de andere kennen. Deze methode voor het vinden van E_A of E_D is de handigste voor het bepalen van bindingsenergieën. Een ander gegeven dat men aan de intensiteitsverdeling in het spectrum en de energie als functie van r kan ontlenen, is het antwoord op de vraag of de twee verontreinigingen (donors en acceptors) verschillende of identieke soorten roosterplaatsen bezetten. Deze extra kennis betekent, naast het gegeven van de verdeling van de verontreinigingen over de bepaalde roosterplaatsen, natuurlijk ook dat we weten welke atomen (Ga of P) zijn vervangen. De langs deze weg gevonden bindingsenergieën zijn voor een aantal verontreinigingen verzameld in tabel V van het in [14] aangehaalde artikel.

Het bovenstaande leert ons dat het onderzoek van paarspectra, gecombineerd met de kennis van de aanwezige verontreinigingen, fysische details verschaft van het recombinatiemechanisme waarbij zichtbaar licht wordt opgewekt, en tevens van de chemische aard van de donors en acceptors.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

III-B. De diepe donor zuurstof

Ofschoon de groene luminescentie, veroorzaakt door paaremissie van ondiepe donor- en acceptorniveaus bij een temperatuur van 1,6 K, een rendement heeft van 100%, verdwijnt deze emissie bij het bereiken van de kamertemperatuur. Dit komt doordat de ladingdragers bij deze temperatuur door thermische excitatie worden vrijgemaakt en doordat ze recombineren via niet-stralende centra. Wanneer er echter sprake is van een *diep* donor- of acceptor-niveau, kan de paaremissie ook bij kamertemperatuur nog een redelijk goed rendement hebben. Een voorbeeld van dit soort luminescentie vinden we bij het diepe donorniveau van zuurstof. Dit niveau ligt bij 1,6 K ongeveer 0,90 eV beneden de geleidingsband. Bij kamertemperatuur is dit energieverschil maar weinig kleiner: $(E_D)_o \approx 0{,}83$ eV, een bedrag dat voor $T = 300$ K ongeveer overeenkomt met $32\,kT$. Dit laatste betekent dat een elektron dat gevangen zit op een zuurstofdonor maar een zeer kleine kans heeft bij 300 K langs thermische weg vrij te komen. Een dergelijk elektron op een diepe zuurstofdonor kan bij lage temperatuur recombineren met een gat, gebonden aan een acceptor. Het paarspectrum dat optreedt met Zn als acceptor is afgebeeld in *fig. 5*. De emissie ligt in het gebied van het nabije infrarood. Uit een onderzoek van dit O-Zn-donor-acceptor-paarspectrum heeft men afgeleid [22] dat $(E_D)_o = 0{,}895 \pm 0{,}001$ eV.

Eigenlijk verwacht men dat emissie in het infrarood, afkomstig van het paar O(donor)-Zn(acceptor), bij kamertemperatuur zou ontbreken, daar dan de meeste Zn-acceptors geïoniseerd zijn. Bij kamertemperatuur is er al een aanzienlijk aantal vrije gaten in de valentieband, ongeveer gelijk aan het aantal zinkatomen. Deze gaten hebben een golffunctie die op een vlakke golf lijkt, en die derhalve met de in de diepe zuurstofdonorniveaus gevangen elektronen in wisselwerking kan treden. Deze wisselwerking resulteert in een recombinatie van een gebonden elektron met een vrij gat [23]. Hoe het spectrum nu precies verandert wanneer men van een donor-acceptor-recombinatie overgaat naar zo'n recombinatie van een gebonden elektron met een vrij gat, ziet men in *fig. 6*. De piek in de emissiecurve bij 1,35 eV verschuift naar een hogere waarde van de ener-

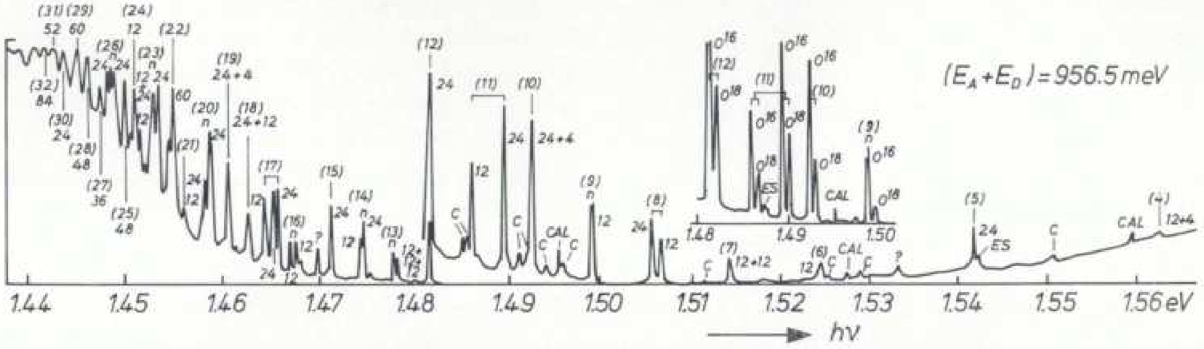


Fig. 5. Het spectrum van de infrarode luminescentie van GaP met (Zn,O)-dotering, afkomstig van paaremissie waarbij een diep zuurstofdonorniveau en een ondiep zinkacceptorniveau zijn betrokken, gemeten bij 1,6 K. De tussen haakjes geplaatste getallen zijn de schilnummers en de andere getallen geven de zg. paarontaarding aan, waarmee het aantal niet-equivalente paarposities binnen een gegeven schil in rekening wordt gebracht. De inzet laat verschuivingen zien die in het spectrum optreden wanneer men een deel van de ^{16}O-atomen door ^{18}O-atomen vervangt. (Ontleend aan Dean c.s. [22].)

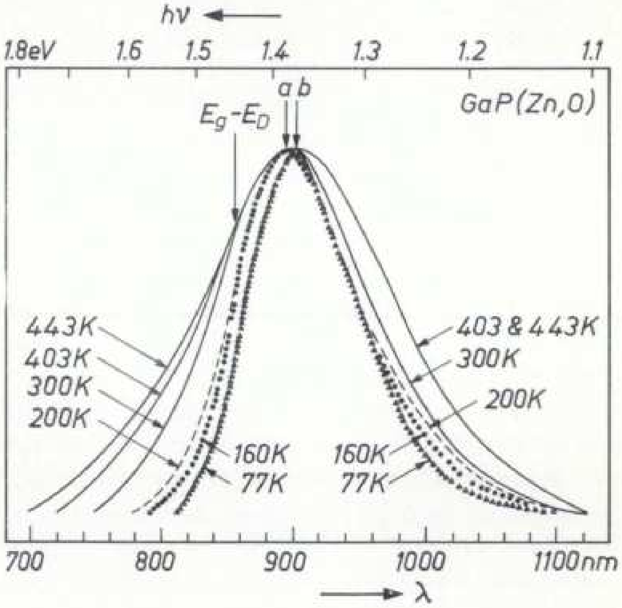


Fig. 6. Variatie met de temperatuur van het luminescentiespectrum (in het infrarood) van GaP met (Zn,O)-dotering. De donor-acceptor-paaremissie gaat geleidelijk over in emissie door recombinatie van gebonden elektronen met vrije gaten (zie ook fig. 7*b*). De twee pijltjes bij het intensiteitsmaximum geven de verschuiving van het maximum aan die optreedt bij een temperatuurdaling van 160 K (zwarte stippen; piek bij *a*) naar 77 K (zwarte driehoekjes; piek bij *b*). De verschuiving van het maximum gaat in een richting die tegengesteld is aan dat wat de bandafstand doet bij dezelfde temperatuurverandering.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

gie als de temperatuur stijgt. Dit kan eenvoudig worden verklaard aan de hand van de volgende uitdrukkingen voor de DA-(donor-acceptor)- en de BF-(gebonden elektron-vrij gat)recombinatie:

$$E_{h\nu}^{DA} = E_g - (E_A + E_D) + e^2/\varepsilon r - E_f \qquad (2)$$

en

$$E_{h\nu}^{BF} = E_g - E_D - E_f, \qquad (3)$$

waarin $E_{h\nu}^{DA}$ en $E_{h\nu}^{BF}$ de energiewaarden zijn van de waargenomen intensiteitspieken; E_f is de energie van de fononen die afkomstig is van het in het diepe donorniveau stevig vastzittende elektron en daarom aan beide situaties op gelijke wijze bijdraagt. In het O-Zn-systeem is E_A gelijk aan 64 meV, terwijl $e^2/\varepsilon r$ ongeveer 40 meV bedraagt; dit betekent dat we verwachten dat $E_{h\nu}^{BF}$ ongeveer 24 meV verschoven zal zijn in de richting van hogere energiewaarden. De waargenomen verschuiving (fig. 6) komt daarmee redelijk overeen. Men mag zeggen dat bij stijgende temperatuur het recombinatiemechanisme bij 77 K van overwegend DA al volledig BF is geworden. Zoals we hieronder nog zullen uitleggen, is dit laatste resultaat ook in overeenstemming met metingen van de levensduur van elektronen die zich in het diepe donorniveau bevinden.

Als we de situatie enigszins simplificeren, kan gezegd worden dat het uitsterven van luminescentie die ontstaat bij recombinatie van een elektron met een gat, een eenvoudige exponentiële wet volgt: $I(t) = I(0)$ exp $(-t/\tau_D)$, waarin $I(t)$ de intensiteit van het luminescentielicht is op een gegeven tijdstip t en τ_D de tijdconstante. Voor $t = \tau_D$ geldt $I(t) = I(0)/e$; de gemeten waarde voor τ_D noemt men dan ook wel de 1/e-tijd. In het geval van de donor-acceptor-paaremissie hangt de recombinatie af van de onderlinge afstand in het rooster, zodat de snelheid van het uitsterven van de luminescentie voor verschillende posities in het rooster verschillend is. De waargenomen snelheid ontstaat uit een som van vele exponentiële vervalfuncties, wat een niet-exponentieel verloop van het uitsterven van de luminescentie tot gevolg heeft; voor het gemak werkt men echter toch met één karakteristieke tijdconstante van de 1/e-soort. In het geval van recombinatie van een gebonden elektron met een vrij gat verwachten we een zuiver exponentiële tijdsafhankelijkheid. Door het uitsterven van de luminescentie nauwkeurig te onderzoeken, kan men dus uitmaken of het gaat om een donor-acceptor-paaremissie of om de recombinatie van een gebonden elektron met een vrij gat. Op deze wijze hebben we ook kunnen bevestigen dat bij temperaturen hoger dan 150 K de infrarode emissie van 1,35 eV afkomstig is van de recombinatie van een gebonden elektron met een vrij gat [23].

De emissie van infrarood bij kamertemperatuur moge dan veroorzaakt worden door de recombinatie van een elektron in een diep donorniveau met een vrij gat, toch is het rendement ervan niet groot. Kwalitatief kan men dit verklaren door te kijken naar het proces waarmee elektronen uit de geleidingsband in de grondtoestand van de zuurstofdonor worden gevangen, en dan in aanmerking te nemen dat de concentratie van deze donors gering is. Het zuurstofdonorniveau ligt zo diep dat wanneer de donor een elektron uit de geleidingsband vangt, dit elektron een hoeveelheid energie moet verliezen gelijk aan die van 17 zg. optische fononen; de energie van één optisch fonon is 50,1 meV. Dit wijst er reeds op dat de kans elektronen in een dergelijk diep niveau te vangen, uiterst klein is. Er is echter nog een andere mogelijkheid: het elektron kan gevangen worden in een aangeslagen toestand van de donor, waarna dan een overgang naar de grondtoestand onder emissie van straling plaatsvindt. Dit is inderdaad door Dean en Henry waargenomen bij zuurstof [24]. De overgang wordt wel als inwendige-vangstluminescentie in de zuurstofdonor betiteld; zie ook *fig. 7*. Meting van de uitsterftijd van deze luminescentie geeft voor lage temperaturen ongeveer 20 μs [13]. Indien we veronderstellen dat deze uitsterftijd ook bij stijgende temperatuur constant blijft, mogen we verwachten dat dan het aantal aangeslagen zuurstofdonors snel zal verzadigen als een toenemend aantal elektronen wordt geïnjecteerd in de geleidingsband. Is de verzadigde toestand bereikt, dan zal de effectieve vangtijd voor elektronen ten minste 20 μs zijn. Indien dit juist is, zou men het quantumrendement van de infrarode emissie kunnen schatten door het quantumrendement van het rood licht uitzendende zink-zuurstof-complex te bestuderen en de vangtijden te vergelijken; voor het Zn-O-complex is die ongeveer 20 ns. Uit de verhouding van deze twee tijden concluderen we dat de infrarode emissie ongeveer 10^3 maal zo zwak zal zijn als de rode, zodat het quantumrendement voor de infrarode emissie in de orde van grootte van 10^{-4} komt te liggen.

Verder wordt het lage rendement van de infrarode emissie veroorzaakt door het geringe aantal atomen van de substituent (O). Nieuwe gegevens van Kukimoto c.s. [25], verkregen uit fotocapaciteitsmetingen, wijzen erop dat het aantal vrije zuurstofdonors in *P*-GaP slechts ongeveer $2 \times 10^{16}/cm^3$ bedraagt.

Ofschoon de zuurstofdonor als licht emitterend centrum van minder belang is, zullen we in het vervolg van dit artikel nog zien dat er dank zij deze donor wèl zeer interessante andere stralende centra kunnen ontstaan, en ook niet-stralende, in samenhang met hetzij bepaalde verontreinigingen hetzij roosterfouten (vacatures). Het zal dan ook duidelijk zijn dat de chemie van het inbouwen van zuurstof van groot belang is voor het begrijpen van het ontstaan van licht emitterende centra.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

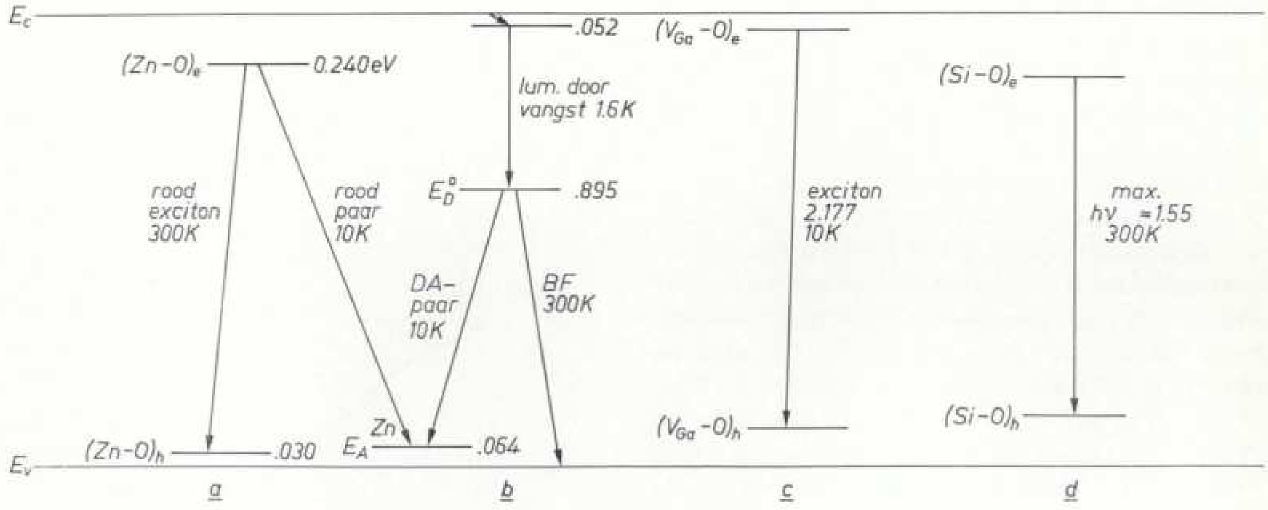


Fig. 7. Schematische voorstelling van een aantal stralingsovergangen waarbij een geïsoleerde zuurstofdonor of een zuurstof bevattend complex betrokken is. *a*) De overgang van een exciton dat gebonden is aan een Zn-O-complex van het naaste-buurtype. $(Zn\text{-}O)_e$ en $(Zn\text{-}O)_h$ zijn de bindingsenergie van het elektron en het gat. Bij lage temperaturen als 10 K kan het elektron ook recombineren met een gat bij een zinkacceptor (E_A^{Zn}), waarbij rood licht met een „paar-achtig" spectrum wordt uitgezonden. *b*) „Vangstluminescentie" in het infrarood (≈ 0,84 eV), door overgang van een aangeslagen toestand van de zuurstofdonor naar diens grondtoestand E_D^O, bij 1,6 K. Vanuit de grondtoestand kan het elektron recombineren hetzij met een gat op een zink-acceptor, hetzij met een gat in de valentieband. De eerste recombinatie leidt tot de donor-acceptor-paaremissie (*DA*), de tweede tot de emissie horend bij de recombinatie van een gebonden elektron en een vrij gat (*BF*). *c*) De overgang van een exciton, gebonden aan een galliumvacature/zuurstofdonor-complex ($V_{Ga}\text{-}O_P$), leidend tot emissie in het oranje deel van het spectrum. *d*) De overgang bij $Si_{Ga}^+\text{-}O_P^+$, een naaste-buurcomplex, geschiedt overwegend stralingsloos (zie hoofdstuk VIII), eventuele emissie ligt in het infrarood.

IV. Iso-elektronische centra; stikstof

In het vorige hoofdstuk zijn de luminescentie-eigenschappen van een zuurstofdonor beschreven, die in GaP een diep liggende vangplaats voor elektronen is. In dit hoofdstuk gaan we de luminescentie behandelen die afkomstig is van een ander soort verontreinigingen, de reeds genoemde iso-elektronische centra. Zij vertonen wel enige eigenschappen van de zojuist behandelde diep liggende niveaus, maar de bindingsenergie ervan is lang niet zo groot. Deze iso-elektronische centra zijn in GaP ontdekt in 1965 [26], en ze zijn van doorslaggevende betekenis omdat ze de snelheid van het recombinatieproces in een halfgeleider met indirecte kleinste bandafstand sterk kunnen vergroten. We zullen hoofdzakelijk nagaan hoe deze iso-elektronische centra in staat zijn het luminescentierendement te verhogen. Het volgende heeft betrekking op GaP met een stikstofdotering; stikstof geeft namelijk aanleiding tot het ontstaan van het eenvoudigste iso-elektronische centrum, en is ook verantwoordelijk voor de groene luminescentie in GaP-LED's met hoog rendement [27]. Zo'n iso-elektronisch centrum is in feite een iso-elektronische substituent (dat is een verontreinigingsatoom dat een in dezelfde kolom van het periodieke systeem thuis horend atoom van het kristalrooster vervangt) die hetzij een elektron, hetzij een gat kan binden. Zodra nu één ladingdrager gebonden is, heeft het centrum een lading; daardoor zal een vrije ladingdrager van het tegengestelde teken gemakkelijk door Coulomb-aantrekking worden gevangen. Dit betekent dat nu aan het centrum een exciton gebonden is. Als voorbeeld: stikstofatomen die fosforplaatsen in GaP innemen en die dan een elektron kunnen binden. Kwalitatief valt dit wel te begrijpen: men mag verwachten dat stikstof elektronen sterker aantrekt dan fosfor, omdat de positieve atoomkern bij stikstof nu eenmaal minder goed afgeschermd wordt. De situatie is dan deze dat het elektron gemiddeld ongeveer op de plaats van het centrale stikstofatoom zit en dat het (later) aangetrokken gat er in een Coulomb-baan omheen blijft lopen; het geheel lijkt wel wat op een acceptor en men spreekt dan ook van een iso-elektronische acceptor [28]. Zou het gaan om een iso-elektronisch atoom van een zwaarder element dan fosfor, dan mag men verwachten dat een vrij gat min of meer centraal gebonden raakt; een later door de Coulomb-kracht gevangen elektron blijft er dan in een baan omheen lopen, wat resulteert in een iso-elektronische donor. Dit geval doet zich voor wanneer bismut het vervangende atoom is. Het element arseen daarentegen, leidt niet tot zo'n gebonden toestand in de verboden zone van GaP [13].

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Steeds wanneer een deeltje gebonden wordt door een verontreinigingscentrum betekent dit een bepaalde lokalisering van het deeltje in de „echte" ruimte. Aangezien de radiusvector $\boldsymbol{r}$ uit de echte ruimte en de golfvector $\boldsymbol{k}$ van de $\boldsymbol{k}$-ruimte complementaire grootheden zijn, zal de golffunctie in de $\boldsymbol{k}$-ruimte een des te groter gebied beslaan naarmate het desbetreffende deeltje sterker gebonden is. In *fig. 8* ziet men schematisch geïllustreerd hoe groot dit effect voor een elektron kan zijn. De mate waarin de golffunctie zich uitbreidt, hangt ten nauwste samen met het soort binding. Bij een „normale" donor of acceptor, waar het een Coulomb-binding betreft, is de ruimtelijke uitbreiding van de golffunctie zeer gering [13] zoals geïllustreerd is voor een gat in fig. 8. De oorzaak hiervan is de lange dracht van de Coulomb-kracht, de golffunctie vertoont daardoor in de werkelijke ruimte geen abrupte veranderingen. Er bestaan echter krachten met korte dracht, die niets te maken hebben met de Coulomb-kracht. Een dergelijk geval doet zich voor bij N in GaP; men denkt dat de gebonden toestand daar hoofdzakelijk te danken is aan het grote verschil in elektronegatieviteit tussen het N-atoom en het erdoor vervangen P-atoom. Zo'n pseudo-potentiaalverschil op zichzelf zou al leiden tot een bindingsenergie voor een elektron van de orde van 1 eV. De hydrostatische vervorming van het GaP-rooster rond het N-atoom zal deze bindingsenergie verminderen. Het resultaat hiervan is dat het stikstofatoom een elektron kan binden met een energie van ongeveer 10 meV [28-30]. De bindende kracht is slechts merkbaar over één of twee roosterafstanden. Vanwege de sterk gelokaliseerde vervormingen die deze veroorzaakt in dat deel van de golffunctie van het elektron dat geldt voor de naaste omgeving van het stikstofatoom, zal de golffunctie zich door de gehele $\boldsymbol{k}$-ruimte uitbreiden. Bovendien vertoont deze uitbreiding van de golffunctie, door de aanwezigheid van een nevenminimum in de geleidingsband bij $k = 0$ (in het centrum van de Brillouin-zone dus, zie fig. 2), een sterke concentratie rond $k = 0$. Dit correspondeert met de k-waarde van het gat (vgl. fig. 8), en derhalve kan recombinatie gemakkelijk optreden. Dit verklaart waarom het iso-elektronische centrum stikstof stralende recombinatie dicht bij de bandkant (in het groene deel van het spectrum) zo sterk bevordert (zie *fig. 9*, en ook fig. 2). Het spectrum laat bij een temperatuur van 1,6 K een scherpe „fononloze" lijn van 2,3171 eV zien (de A-lijn); fig 9*a* toont dat er ook sterke lijnen met „fononassistentie" uitgezonden worden. Aangezien de bindingsenergie bij zo'n iso-elektronisch centrum bijzonder gevoelig is voor deformaties van de centrale roostercel, zullen de zg. optische roostertrillingen een grote bijdrage aan de fononassistentie leveren. Het spectrum in fig. 9*b* is dat van een groen licht emitterende diode, bij 300 K [27]. De hoog-

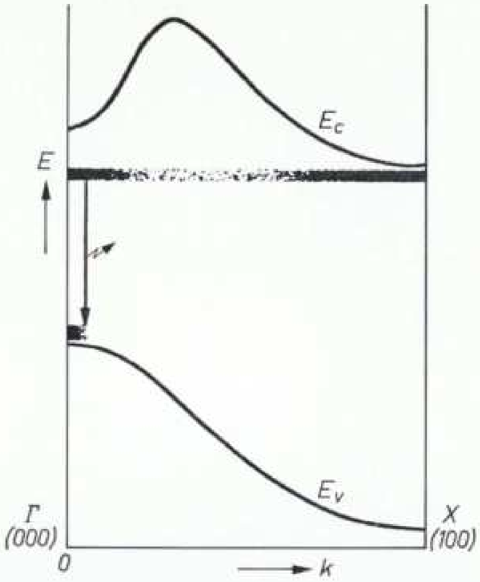


Fig. 8. Voorstelling van de waarschijnlijkheidsdichtheid, door middel van de variabele zwarting, voor een elektron gebonden aan een iso-elektronisch centrum (N in GaP) dat in de buurt van de geleidingsband (minimale energie E_c) ligt, en voor een gat dat gebonden is aan een acceptor op een ondiep niveau in de buurt van de valentieband (maximale energie E_v). De waarschijnlijkheidsdichtheid is uitgezet als functie van de golfvector $\boldsymbol{k}$. Het feit dat deze waarschijnlijkheid in het geval van het elektron in de gehele $\boldsymbol{k}$-ruimte van nul verschilt, is de oorzaak van het goede rendement van de luminescentie in GaP met N-dotering. (Ontleend aan Thomas [15].)

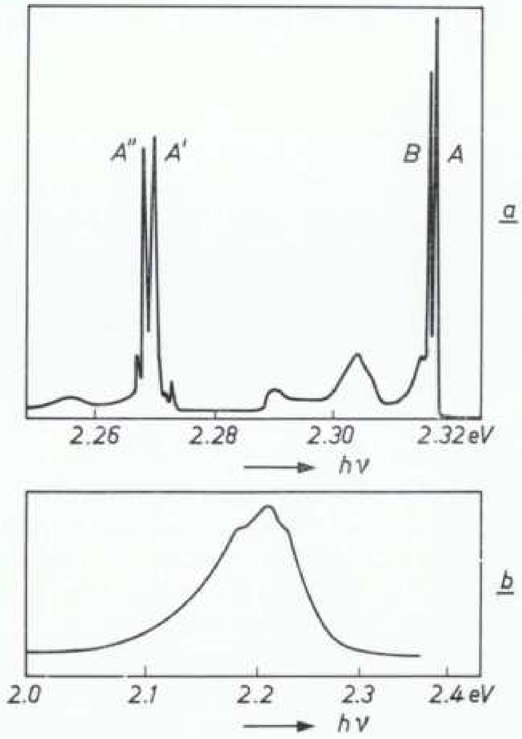


Fig. 9. Luminescentiespectra van GaP met N-dotering. *a*) Het fotoluminescentiespectrum voor excitonen gebonden aan N, gemeten bij 1,6 K. Er zijn twee fononloze lijnen (*A* en *B*) te zien en hun satellieten, *A'* en *A''*, die een gevolg zijn van overgangen met fononassistentie. Deze satellieten hebben een aanzienlijke intensiteit; zij hebben voornamelijk te maken met optische fononen, en liggen bij ongeveer 2,27 eV. *b*) Een luminescentiespectrum bij 300 K van een GaP(N)-diode. De energie van het maximum correspondeert met de satelliet *A'* van de lijn *A*. (Ontleend aan Logan, White en Wiegmann [27].)

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

ste intensiteit treedt op bij de energie van de A-lijn verminderd met die van één optisch fonon (*LO*). Het maximale uitwendige quantumrendement dat tot dusver met groene diodes is behaald, bedraagt 0,6% [27], waarmee een lichtstroomrendement van 3,6 lm/W overeenkomt.

Een opmerkelijk feit is nog dat het iso-elektronische centrum bestaande uit een bismutverontreiniging, bij 300 K *niet* tot luminescentie met een bevredigend rendement leidt, ofschoon dit centrum dieper ligt (de bindingsenergie voor een gat bedraagt daar 0,04 eV). Het vervangen van stikstof door bismut levert blijkbaar voor de excitonen een veel kleinere overgangswaarschijnlijkheid op. Dit is verklaarbaar uit het feit dat bij bismut juist het *gat* gebonden is, door een kracht met korte dracht; dit laatste betekent weer een gelijkmatige verspreiding van de golffunctie over de Brillouin-zone. De bandstructuur van GaP vertoont geen maximum in de valentieband op het punt waar de geleidingsband zijn minimum heeft. Integendeel, op dit punt ligt de rand van de valentieband een aantal elektronvolts lager dan het maximum, dat bij $k = 0$ (centrum van de zone) optreedt. Het gevolg van deze situatie is dat er bij bismutcentra veel minder gelegenheid bestaat voor een rechtstreekse recombinatie van een elektron en een gat dan bij stikstofcentra.

Men merke nog op dat bij geen van beide iso-elektronische centra de zg. impulsbehoudende fononen overwegen in de gevonden fononzijbanden, hoewel nog dikwijls wordt gedacht dat optische overgangen bij halfgeleiders met niet-verticale kleinste bandafstand hun sterkte voornamelijk ontlenen aan deze fononen.

V. Naaste-buurcomplexen

Bij de opwekking van rood licht door GaP met een (Zn,O)-dotering wordt gebruik gemaakt van de beide eigenschappen die men voor het verkrijgen van een goed rendement nodig heeft, namelijk (a) het energieniveau van het centrum ligt zo diep dat gebonden deeltjes langs thermische weg niet of nauwelijks kunnen worden vrijgemaakt, en (b) de lokalisering van het erbij betrokken elektron-gatpaar geschiedt door het vormen van een exciton dat door een iso-elektronisch centrum gebonden wordt. Het is interessant om na te gaan hoe dit alles wordt verwezenlijkt. Zoals we in hoofdstuk III reeds zagen, ligt het zuurstofdonorniveau bij 300 K ongeveer 0,83 eV beneden de geleidingsband, en door deze grote diepte kan slechts nutteloze straling in het nabije infrarood worden opgewekt. De grote diepte ontstaat niet door Coulomb-aantrekking, maar door hetzelfde soort krachten met korte dracht als bij de besproken stikstofverontreiniging; de golffunctie van het gebonden elektron is dus voornamelijk geconcentreerd in de buurt van de centrale cel. Bijgevolg ontstaat weer een gelijkmatige verdeling van de golffunctie over de *k*-ruimte.

Door nu in de directe omgeving een acceptor met tegengestelde lading aan te brengen, kan men de waarde van de bindingsenergie zodanig veranderen dat er bruikbare straling wordt opgewekt in plaats van infrarood. Dit is het geval wanneer een Ga-atoom op een naaste-buurplaats vervangen wordt door een Zn-atoom. De negatieve lading van de Zn_{Ga}-acceptor reduceert de bindingsenergie van het elektron tot 0,24 eV en er ontstaat een molecuulachtig iso-elektronisch centrum (complex) [31] [32]. Dit oorspronkelijk neutrale Zn_{Ga}^{-}-O_{P}^{+}-centrum zal, nadat het een elektron gevangen heeft, een gat in een Coulomb-baan kunnen binden. Recombinatie van dit gat met het elektron resulteert dan in een luminescentie met zeer goed rendement. In de literatuur zijn uitwendige quantumrendementen van ongeveer 85% bij 77 K en ongeveer 21% bij 300 K gerapporteerd [33] [34], althans voor *foto*luminescentie. Het hoogste uitwendige quantumrendement voor rode LED's dat men vermeld vindt, bedraagt 12% [18]. We mogen echter verwachten dat in de toekomst de quantumrendementen bij LED's (elektroluminescentie) ongeveer gelijk zullen zijn aan de waarden die gelden voor fotoluminescentie, omdat men dan de vorming van de *P*-*N*-overgang ongetwijfeld beter zal begrijpen en beheersen.

V-A. Identificatie van licht emitterende centra

Dat de rode luminescentie geïdentificeerd moet worden als de annihilatie van een gebonden exciton bij een Zn-O-naaste-buurcomplex heeft een interessante voorgeschiedenis. Allen c.s. [5] hebben in 1963 laten zien dat men met door legering gevonden *P*-*N*-overgangen, gemaakt op uit een oplossing gegroeide plaatjes GaP met (Zn,O)-dotering, LED's kan verkrijgen met rendementen van ongeveer 0,01 tot 0,1%. Grimmeiss en Scholz van Philips Forschungslaboratorium Aachen [6] rapporteerden in 1964 een rendement van niet minder dan 1,5%, doch de reproduceerbaarheid was onbevredigend. In 1966 kwamen Lorenz en Pilkuhn [35] met de meest belovende methode voor het maken van *P*-*N*-overgangen in GaP; hun methode (LPE genoemd naar „liquid-phase epitaxy") is een uitbreiding van een methode die Nelson als eerste in 1963 ontwikkeld heeft voor het maken van overgangen in GaAs [36].

De vorming van *P*-*N*-overgangen met behulp van LPE bestaat uit het in aanraking brengen van een kristallijn GaP-substraat (b.v. van geleidingstype *N*, met Te-dotering) met een oplossing van GaP, Ga, en geschikte doteringen, die in het gekozen voorbeeld Zn en Ga_2O_3 kunnen zijn (wat leidt tot een laag van geleidingstype *P*). Men doet dat door kantelen van de betrokken oven. Vervolgens wordt, vanaf 1100 °C, langzaam gekoeld (ca.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

1-5 °C/min); dan groeit in een periode van een aantal uren een dunne *P*-laag (50-100 μm) op het *N*-substraat. Tot nu toe zijn de beste resultaten verkregen met *P*-lagen op een *N*-substraat.

In de laatste jaren zijn grote eenkristallen van GaP ter beschikking gekomen [37], die gemaakt zijn met behulp van de Czochralski-methode. Men heeft daar rode LED's mee vervaardigd door middel van een dubbel LPE-procédé. Het komt erop neer dat men eerst op een N^+-substraat met LPE een *N*-laag met Te-dotering ($\approx 7 \times 10^{17}\,cm^{-3}$) laat groeien, waarna men daarop een *P*-laag met (Zn,O)-dotering ($\approx 3 \times 10^{17}\,cm^{-3}$) aanbrengt. Met een op deze wijze vervaardigde LED is het uitwendige quantumrendement van 12% bereikt dat hierboven genoemd werd [18].

Oorspronkelijk leidde het fysisch onderzoek van de stralingsprocessen in GaP met (Zn,O)-dotering tot de opvatting dat de luminescentie van het rode licht veroorzaakt zou worden door paaremissie waarbij op (verre) afstand gelegen DA-paren van zuurstof en zink betrokken zouden zijn [38]. Men meende dat het zuurstofdonorniveau op een diepte van ca. 0,47 eV zou liggen, en overeenkomstig dit beeld schreef men het goede rendement toe aan de sterke binding van het elektron. In 1967 kon worden bewezen dat deze opvattingen niet juist zijn [31] [32]. Toen men namelijk in plaats van Zn het chemisch daarmee overeenkomende element Cd gebruikte, bleek een volledig ander spectrum op te treden (*fig. 10a*). Bij een temperatuur van 16 K vertoont dit spectrum een scherpe lijn bij 1,907 eV, de zg. A-lijn, en een bijbehorende fijnstructuur als gevolg van fononherhalingen (phonon replicas). Fig. 10*b* leert dat deze lijnen in het spectrum ontbreken wanneer we het GaP gewoon met Zn en O doteren. De scherpe vorm van de fononloze A-lijn maakt het mogelijk nauwkeurige experimenten te doen. Uit deze onderzoekingen, die we in een viertal punten hieronder zullen beschrijven, volgt dat de emissie van rood licht bij GaP met een (Cd,O)-dotering afkomstig is van recombinatie van een exciton, gebonden aan een naaste-buurcomplex Cd_{Ga}^{-}-O_P^{+}.

1. Een bijzonder vruchtbare manier om vast te stellen welke chemische verontreinigingen de emissie veroorzaken, is het doen van een „isotopenproef". We doteren dan met een isotoop die in het natuurlijke isotopenmengsel van het betrokken element *niet* voorkomt. Bij gebruik van het isotoop ^{18}O in plaats van het normale ^{16}O, neemt men waar dat de scherpe A-lijn verschuift naar een 0,7 meV hogere energie [31]. Wanneer men echter op overeenkomstige wijze ^{114}Cd toevoegt en niet ^{110}Cd, dan verschuift de A-lijn niet significant, maar zijn fononsatellieten doen dit wel [32]. Zonder uitvoerig in te gaan op de moeilijke vraag naar het waarom van deze verschuivingen, mag men buiten twijfel als bewezen achten dat Cd en O bij de emissie betrokken zijn.

2. Uit magneto-optische experimenten heeft men de oriëntatie van de atomen Cd en O ten opzichte van elkaar kunnen bepalen [32]. Door afkoeling tot 1,6 K komt een nieuwe lijn naar voren met een 2,3 meV hogere energie dan de A-lijn. De oorzaak hiervan is dat de spins van het gat en het elektron die gevangen zijn in het bij de emissie betrokken centrum, bij de afkoeling overgaan van de toestand waarin ze antiparallel gericht zijn naar die waarin ze parallel staan. De stralingsovergang die correspondeert met dit lagere energieniveau noemt men de B-lijn. Bij deze B-lijn blijkt een sterk richtingsafhankelijke Zeeman-splitsing op te tre-

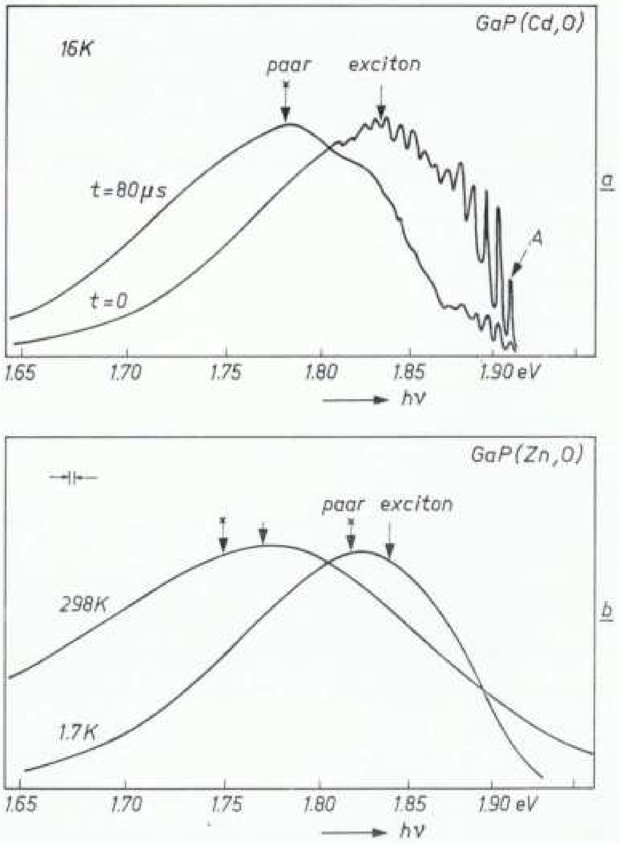


Fig. 10. De luminescentiespectra van GaP met (Cd,O)- of (Zn,O)-dotering; de spectra ontstaan door overgangen van gebonden excitonen en overgangen van een paarachtig type, bij naastebuurcomplexen (Cd-O of Zn-O). De pijltjes geven de plaats van de maxima van de beide soorten spectra aan. *a*) Het spectrum van het Cd-O-complex op het tijdstip dat de excitatie stopt ($t = 0$) en 80 μs later. Op $t = 0$ domineert een spectrumstructuur die te maken heeft met het excitonverval. Men ziet een fononloze lijn *A* alsmede lijnen die te maken hebben met sterk gelokaliseerde trillingen of met roosterfononen. Op $t = 80$ μs is er een volledig nieuw spectrum verschenen dat correspondeert met het recombineren van een elektron bij een Cd-O-complex met een gat op een Cd-acceptor. *b*) Het spectrum van het rode licht dat GaP met (Zn,O)-dotering uitzendt, bij twee temperaturen. Ofschoon fijnstructuur niet te zien is, blijken toch de (Zn,O)-dotering en de (Cd,O)-dotering tot emissies met ongeveer dezelfde eigenschappen te leiden. In de linkerbovenhoek is het scheidend vermogen van de energiemeting aangegeven (zie Cuthbert c.s. [39]).

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

den. De hieruit gevonden gegevens stemmen overeen met een model waarin de atomen Cd en O in de [111]-richtingen staan. Van deze richtingen bestaan er vier in het kristal. Bij een magneetveld in de [100]-richting zijn de vier richtingen [111] gelijkwaardig. In die situatie neemt men inderdaad de eenvoudigste Zeeman-splitsing waar. Omdat in het rooster de atomen Ga en P ook langs de [111]-richtingen liggen, kan men gerust concluderen dat Cd een Ga-plaats bezet en O een ernaast gelegen P-plaats. Het in hoofdstuk III beschreven discrete spectrum van paarlijnen, dat te maken heeft met geïsoleerde O-donors en Zn- of Cd-acceptors, bevestigt de gegeven interpretatie.

3. De aard van het recombinatieproces dat de straling opwekt, kan ook nog bevestigd worden door de uitsterftijd van de luminescentie te bestuderen. Zoals uit het reeds eerder besprokene volgt, zal de luminescentie exponentieel uitsterven wanneer ze alleen door een overgang van een gebonden exciton wordt veroorzaakt. In het geval van spectra, horend bij Cd-O, heeft men vastgesteld dat het uitsterven niet exponentieel verloopt [39]. Verder vindt men twee duidelijk van vorm verschillende spectra wanneer men na het einde van de exciterende impuls op twee verschillende tijdstippen meet (fig. 10*a*). Dit heeft geleid tot de opvatting dat het stevig gebonden elektron van het donorniveau Cd-O onder uitzending van straling kan recombineren, hetzij met een gat dat gebonden is aan het complex Cd-O, hetzij met een gat dat gebonden is aan een geïsoleerde Cd-acceptor. (Voor het analoge systeem Zn-O is dit schematisch geïllustreerd in fig. 7.) Ofschoon de laatstgenoemde emissie dus wel te maken heeft met zich op enige afstand bevindende Cd-acceptors, wijkt het spectrum toch af van de gebruikelijke donor-acceptor-paarspectra. De oorzaak hiervan is dat de donor Cd-O na recombinatie weer neutraal wordt, zodat de Coulomb-term in vergelijking (1) voor de energie van het geëmitteerde foton vervalt. Ook bij meting van de spectra als functie van de tijd vinden we het voorgaande bevestigd. Doordat bij een stralingsovergang van een gebonden exciton het gat *gelokaliseerd* is, zal dit type overgang bij kamertemperatuur sterker zijn dan dat van het paarachtige type.

4. Dat de emissie kan worden toegeschreven aan een *naaste*-buurcomplex, impliceert dat warmtebehandelingen er een aanzienlijke invloed op zullen hebben. De moleculaire binding tussen Cd en O (of Zn en O) in het GaP-rooster raakt in de buurt van 900 °C verbroken; snel afkoelen („afschrikken") zal dan de emissie van rood licht aanzienlijk kunnen verzwakken. Omgekeerd zal een langdurige verwarming bij een lagere temperatuur (600 °C) het aantal complexen moeten doen stijgen. Voor het geval van Zn en O is dit alles inderdaad gevonden [40]. Wij hebben het gebruikt voor een uitvoerig onderzoek van de kinetica van de recombinatieprocessen die tot de emissie van het rode licht leiden [41]. Dit soort onderzoek heeft ons in staat gesteld de onderlinge sterkteverhouding van de diverse concurrerende processen (zowel die met lichtemissie als de stralingsloze) te bepalen. Uit de aard der zaak kan een echt stralingsloos proces niet langs optische weg worden bestudeerd, maar in de laatste jaren heeft de bestudering van de combinatie van de invloed van warmtebehandelingen met de recombinatiekinetica er belangrijke gegevens over verschaft. In de volgende paragraaf komen we hier nog wat nader op terug.

Zojuist hebben we de details besproken van het onderzoek dat gedaan is aan het Cd-O-complex. Voor het Zn-O-complex, dat bruikbaar licht in het rode deel van het spectrum kan emitteren, vindt men analoge eigenschappen, zij het dat de scherpe lijnen ontbreken. In feite is de onzekerheid met betrekking tot de oorspronkelijke interpretatie van de rode emissie door GaP met (Zn,O)-dotering, als conventioneel donor-acceptor-paarspectrum, te wijten aan twee onverwachte eigenschappen van de gevonden spectra; ten eerste zag men geen lijnstructuur, en ten tweede bleek het paarachtige spectrum vergezeld te zijn van het spectrum van een exciton, gebonden aan een Zn-O-complex.

V-B. De recombinatiekinetica bij GaP(Zn,O)

In de laatste jaren is veel onderzoek gedaan aan de kinetica van het stralingsproces, gebaseerd op recombinatie van excitonen die gebonden zijn aan Zn-O-complexen van het naaste-buurtype [41-44]. We zullen hiervan een overzicht geven, en de belangrijkste resultaten beschrijven die deze onderzoekingen hebben opgeleverd.

Voor het eenvoudigste model [41] is de recombinatie van elektronen (de minderheidsladingdragers) met gaten (de meerderheidsladingdragers) in *P*-GaP met (Zn,O)-dotering voorgesteld in *fig. 11*. Aangeslagen elektronen in de geleidingsband kunnen hetzij gevangen worden door het Zn-O-complex, hetzij verdwijnen via andere, ongewenste wegen (vanaf nu betiteld met stralingsloze wegen). De tijdconstanten voor de vangst van elektronen door het Zn-O-complex en het ontsnappen daaruit door thermische emissie zijn respectievelijk τ_{nt} en τ_{tn}. De tijdconstante voor vangst door stralingsloze centra zij voorgesteld door τ_n. Zo'n karakteristieke vangtijd die bij een bepaalde verontreiniging hoort, is omgekeerd evenredig met de concentratie ervan. De tijdconstante voor recombinatie met gaten van elektronen uit het complex, noemen we τ_R als gewenste straling ontstaat en τ_{NR} voor de stralingsloze processen. De gemeten uitsterftijdconstante τ_D van de excitonemissie hangt af van het aantal Zn-O-complexen, en

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

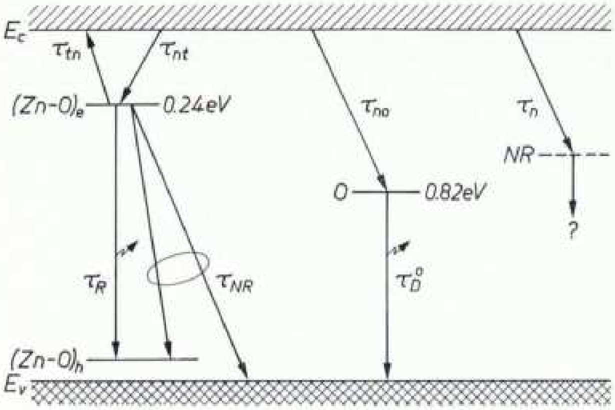


Fig. 11. Schematische voorstelling van het recombinatiemechanisme voor de minderheidsladingdragers (elektronen) in *P*-GaP met (Zn,O)-dotering, bij een temperatuur van 300 K. De tijdconstante voor vangst door het Zn-O-complex is τ_{nt}, de tijdconstante voor thermische emissie uit dit complex is τ_{tn}. De tijdconstante voor vangst door niet-stralende centra is weergegeven door τ_n. De tijdconstante voor recombinatie van een aan het complex gebonden elektron is τ_R voor processen met nuttige straling en τ_{NR} voor de stralingsloze. τ_{nO} en τ_D^O hebben betrekking op vangst, respectievelijk recombinatie bij een zuurstofdonorniveau.

beter rendement, zoals fig. 12 en fig. 13 tonen. Bij de eerste behandeling, het „afschrikken", ziet men τ_n/τ_{nt} dalen, wat een kleinere uitsterftijd en een slechter quantumrendement met zich meebrengt. Op deze wijze kunnen nauwkeurige waarden voor de parameter τ_n/τ_{nt} verkregen worden.

2. Indien men andere vreemde verontreinigingen toevoegt, die de uitsterftijd bekorten, dan mag men concluderen dat de betrokken verontreiniging τ_n/τ_{nt} verandert, mits natuurlijk τ_R en τ_{NR} constant blijven, met andere woorden: mits de concentratie van de vrije gaten niet verandert. Zulk een experiment is gedaan met een chloordotering; inderdaad blijken de uitsterftijd en het quantumrendement kleiner te worden, zoals fig. 12 en fig. 13 leren [45]. Een gedetailleerde bespreking over de niet-stralende centra komt aan de orde in hoofdstuk VIII.

Uit de voorgaande gegevens over uitsterftijd en rendement kan men alleen τ_n/τ_{nt} afleiden. Er zijn verschillende pogingen gedaan [46] [47] om de totale gemiddelde levensduur τ_L van de minderheidsladingdragers te van het aantal niet-stralende centra via de thermische vrijmaking van het elektron dat gebonden is aan het Zn-O-complex (bindingsenergie van het elektron $E_{trap} \approx 0{,}24$ eV bij 300 K); men kan voor de tijdconstante derhalve de volgende uitdrukking geven (zie ook fig. 11):

$$\frac{1}{\tau_D} = \frac{1}{\tau_R} + \frac{1}{\tau_{NR}} + \frac{1}{\tau_{tn}(1 + \tau_n/\tau_{nt})}. \qquad (4)$$

De vangsnelheden $1/\tau_{nt}$ en $1/\tau_n$ zijn respectievelijk evenredig met het aantal licht emitterende centra en het aantal niet-stralende centra. De verhouding van de concentraties van deze twee soorten centra bepaalt het quantumrendement η, waarvoor geldt:

$$\eta = \frac{\tau_D}{\tau_R} \frac{\tau_n/\tau_{nt}}{1 + \tau_n/\tau_{nt}}. \qquad (5)$$

Hierin hangt τ_R (en ook τ_{NR}) af van de concentratie p van de vrije gaten. Door τ_D en η als functie van p uit te zetten, kan men de waarde van τ_n/τ_{nt} schatten (*fig. 12* en *fig. 13*). Voor het verkrijgen van gegevens omtrent de verhouding τ_n/τ_{nt} zijn er twee soorten experimenten gedaan:

1. Kristallen zijn onderzocht vóór en ná de in de vorige paragraaf beschreven warmtebehandelingen: de snelle afkoeling vanaf ca. 900 °C en het stoken bij 600 °C. Bij het stoken is één van de dingen die gebeuren, de vorming van meer Zn-O-complexen, wat resulteert in een toeneming van de verhouding τ_n/τ_{nt}. De hogere waarde van τ_n/τ_{nt} leidt tot een grotere uitsterftijd en een

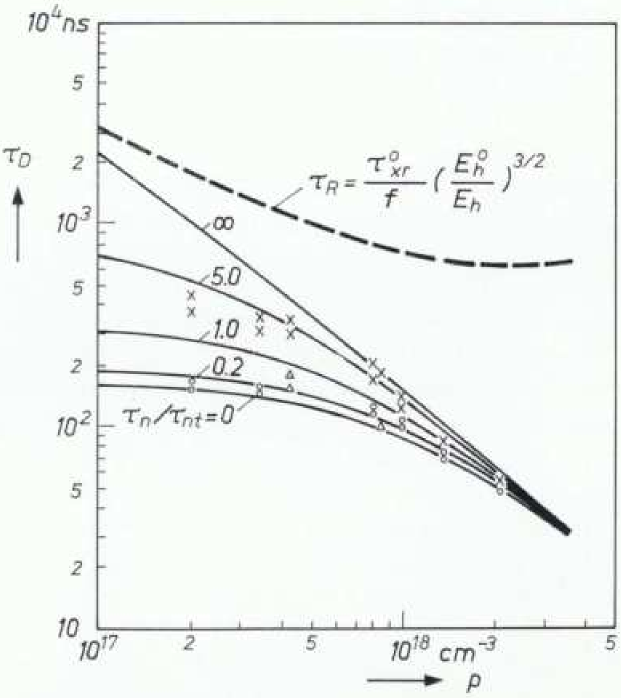


Fig. 12. De voor het uitsterven van luminescentie karakteristieke tijdconstante τ_D, als functie van de concentratie p van vrije gaten in GaP met (Zn,O)-dotering, gemeten bij een temperatuur van 300 K. De getrokken lijnen geven de voorspelde waarden, waarbij als tijdconstante τ_{tn} voor de thermische excitatie (zie ook fig. 11) 161 ns is aangehouden; voor de parameter τ_n/τ_{nt} zijn vijf verschillende waarden gekozen (Jayson c.s. [41]). De kruisjes hebben betrekking op een preparaat dat gedurende een aantal uren bij 600 °C gestookt is, de cirkeltjes op een preparaat dat vanaf een temperatuur van 1000 °C zeer snel is afgekoeld. De eventuele aanwezigheid van galliumchloride bij de warmtebehandeling verslechtert de resultaten (de driehoekjes), wat erop wijst dat onder die omstandigheid de bijdrage van niet-stralende centra stijgt [45].

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

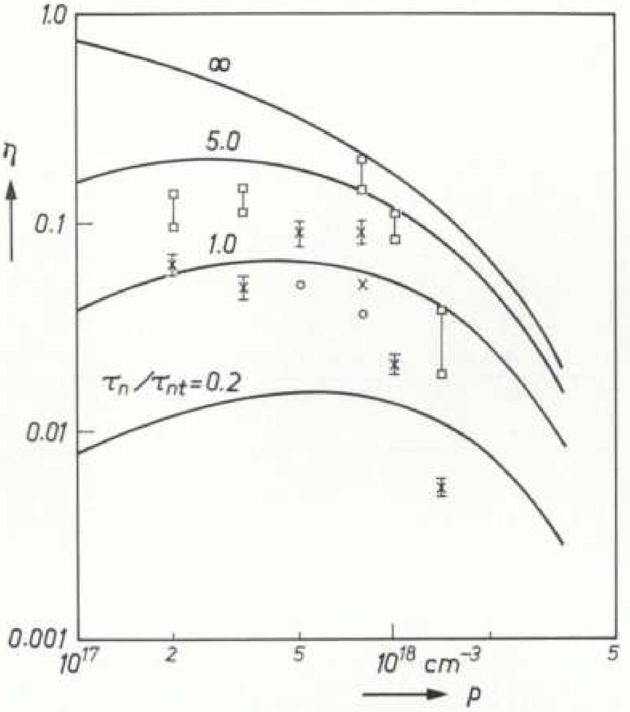


Fig. 13. Het quantumrendement η van de elektronen, in de geleidingsband, als functie van de concentratie p van vrije gaten, en gemeten bij 300 K. De getrokken lijnen geven de voorspelde waarden van het rendement, waarbij als tijdconstante τ_{tn} voor thermische verliezen (zie ook fig. 11) 161 ns is aangehouden; voor de parameter τ_n/τ_{nt} zijn vier verschillende waarden genomen (zie Jayson c.s. [41]). De aanwezigheid van galliumchloride bij de in de tekst beschreven warmtebehandeling van het preparaat doet de waarde van τ_n/τ_{nt} dalen, waardoor ook het rendement zal teruglopen. De vierkantjes geven de voorspelde inwendige quantumrendementen; de kruisjes en cirkeltjes hebben betrekking op gemeten uitwendige quantumrendementen bij preparaten die gestookt zijn in afwezigheid, respectievelijk in aanwezigheid van galliumchloride.

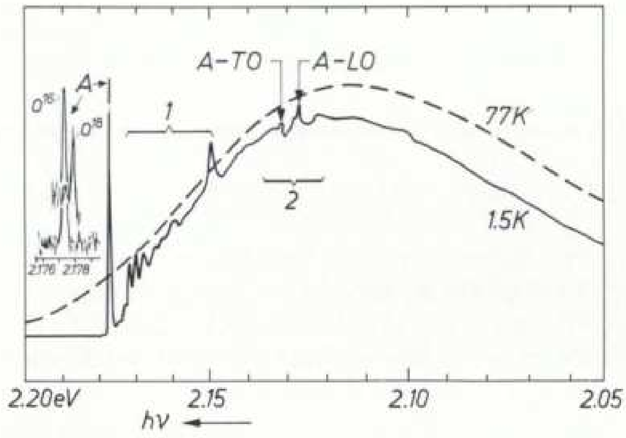


Fig. 14. Het spectrum van de karakteristieke oranje luminescentie (COL), waargenomen bij temperaturen van 1,5 K en 77 K in kristallen die zonder opzettelijke dotering gemaakt zijn met behulp van een waterdamptransportprocédé. De emissie bestaat bij 1,5 K uit een scherpe fononloze lijn (A) en een aantal bijbehorende lijnen die samenhangen met lokaal optredende trillingen (1) en met roostertrillingen (2). De inzet demonstreert de verschuiving van de A-lijn die optreedt door in plaats van ^{16}O de isotoop ^{18}O te gebruiken.

meten; τ_L wordt gedefinieerd door de vergelijking

$$\frac{1}{\tau_L} = \frac{1}{\tau_n} + \frac{1}{\tau_{nt}}. \qquad (6)$$

Wanneer de tijdconstante van de thermische excitatie τ_{tn} veel korter is dan de tijdconstante $(1/\tau_R + 1/\tau_{NR})^{-1}$ voor recombinatie van elektron-gatparen, dan zal de uitsterftijdconstante τ_D van de dan bijna gedoofde luminescentie gelijk zijn aan τ_n. Dit gebeurt al bij ca. 500 °C. Kort geleden heeft men τ_n onder deze omstandigheden kunnen meten [47]. (Zulk een rechtstreekse meting van τ_n geeft informatie over het aantal niet-stralende centra, iets wat op andere wijze door luminescentieonderzoek zeer moeilijk te verkrijgen is.)

Het zal duidelijk zijn dat een groter rendement kan worden verkregen door hetzij het aantal Zn-O-complexen (de centra met nuttige lichtemissie) te vergroten, hetzij het aantal niet-stralende centra te verkleinen. In het volgende hoofdstuk bespreken we hoe deze Zn-O-complexen van het naaste-buurtype zich vormen en hoe men hun aantal zo groot mogelijk kan maken.

VI. Het galliumvacature/zuurstofcomplex

Het is al enige tijd bekend dat het commercieel aantrekkelijk gebleken procédé waarbij groei uit de dampfase plaatsvindt, niet leidt tot GaP met een luminescentie van bevredigend rendement. Recent spectroscopisch onderzoek [48-51] aan dergelijk materiaal heeft hierover waardevolle gegevens verschaft, en het heeft een verder antwoord gegeven op de vraag hoe de naaste-buurcomplexen als Cd-O of Zn-O, die nuttig licht emitteren, in feite gevormd worden. In aanvulling daarop heeft men de vorming van nieuwe nuttige straling emitterende naaste-buurcomplexen zoals Mg-O [52], kunnen verklaren, en ook de aanwezigheid van bepaalde niet-stralende centra [49] [53], waarop we echter pas in de hoofdstukken VII en VIII zullen ingaan.

VI-A. Identificatie

De vondst waarom het hier gaat, is de identificatie van een spectrum, veroorzaakt door de recombinatie van een exciton dat gebonden is aan een naastebuurcomplex van een galliumvacature (V_{Ga}) en een diepliggend zuurstofdonorniveau op de aangrenzende fosforplaats [49] [50] (dit complex zal in het vervolg worden aangeduid met V_{Ga}-O_P). De luminescentie ligt in het oranje deel van het spectrum en het bestaat bij 1,6 K uit een scherpe fononloze lijn van 2,117 eV, in *fig. 14* aangeduid met A. De overige structuur in het spectrum schrijft men toe aan roosterfononen en enige zwakke, met vibratietoestanden samenhangende, wijzen van trilling. Bij toenemende temperatuur blijkt de intensiteit van het spectrum snel af te ne-

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

men. De waargenomen intensiteit van deze karakteristieke oranje luminescentie (in het vervolg met COL aangeduid) is niet groot. De gemeten uitsterftijd ervan bedraagt 110 μs bij een temperatuur van 4 K. De verschillende stappen die hebben geleid tot de identificatie van het COL-spectrum zullen hieronder beschreven worden [*]. De experimenten zijn gedaan met een GaP-kristal dat is gemaakt met behulp van een waterdamp-transportprocédé; dergelijke kristallen vertonen de sterkste COL.

1. Stoken van zulk een kristal in vacuüm gedurende 15 uur bij een temperatuur van 400 °C doet de COL volledig verdwijnen (*fig. 15a*).

2. Indirecte aanwijzingen voor het bestaan van een V_{Ga}-O_P-complex zijn onlangs door Dean [48] verkregen bij de bestudering van het spectrum, afkomstig van een complex Li_{int}-Li_{Ga}-O_P; Li_{int} slaat op een Li-atoom dat een interstitiële plaats bezet. Als in de onder 1 genoemde kristallen bij 400 °C Li werd aangebracht door diffusie gedurende 15 uur, bleek emissie afkomstig van Li_{int}-Li_{Ga}-O_P op te treden (fig. 15*a*). In dit geval is de COL compleet verdwenen doordat de Li-atomen de V_{Ga} in de V_{Ga}-O_P-complexen vervangen, volgens het model van Dean [48]. Wanneer echter bij reeds volgens 1 gestookte kristallen (waarin de COL dus volledig gedoofd is) Li-diffusie wordt toegepast, neemt men een ten minste dertig maal zo zwakke rode luminescentie waar, die afkomstig is van het complex Li_{int}-Li_{Ga}-O_P. Dit is een sterke aanwijzing dat de COL te maken heeft met het V_{Ga}-O_P-complex.

3. De beste bevestiging van de rol die de zuurstof speelt, is weer door een isotopenproef [50], waarbij ^{18}O in plaats van ^{16}O wordt gebruikt. De inzet in fig. 14 toont de verschuiving in het spectrum ten gevolge van de verandering van isotoop; dit is een onweerlegbaar bewijs voor het feit dat zuurstof deel uitmaakt van het centrum dat tot de COL aanleiding geeft.

4. Magneto-optische gegevens [50] wijzen erop dat de stralingsovergang afkomstig is van een gebonden exciton en dat er als eindtoestand een neutrale dubbele acceptor ontstaan is. In die neutrale acceptor zijn de spins van de beide gaten antiparallel. Verder heeft men kunnen nagaan dat de sterke binding (met 160 meV) van het exciton aan V_{Ga}-O_P voor een aanzienlijk deel afkomstig is van het gebonden gat, dat is een andere situatie dan bij de Cd-O- en de Li_{int}-Li_{Ga}-O_P-complexen, waar het gat slechts zwak gebonden is aan het elektron door een Coulomb-kracht, dat wil zeggen een kracht met lange dracht.

VI-B. De rol van het complex V_{Ga}-O_P

Nu bevestigd is dat de COL ontstaat door een recombinatie bij een V_{Ga}-O_P-complex, zullen we in deze paragraaf enkele aspecten van het bestaan van een dergelijk complex bespreken; aspecten, waarvan we geloven dat ze in het algemeen van belang zijn bij de vorming van een aantal complexen met en zonder emissie van nuttig licht in GaP.

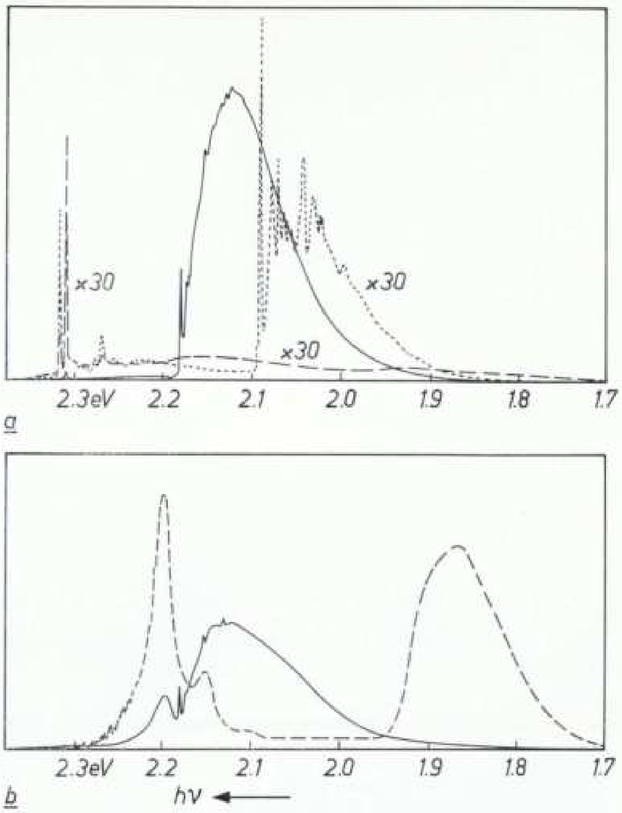


Fig. 15. De invloed van stoken op het foto-luminescentie-spectrum bij preparaten die een sterke karakteristieke emissie in het oranje (COL) vertonen. *a*) Van een ongedoteerd GaP-kristal dat gemaakt is met behulp van een waterdamptransportprocédé, en dat vervolgens gedurende 15 uur bij 400 °C is gestookt in vacuüm (onderbroken lijn) of dat, tijdens het stoken, is verhit in aanwezigheid van Li (streeplijn). Stoken in vacuüm zonder meer doet de intensiteit van de COL aanzienlijk dalen. De Li-diffusie tijdens het stoken leidt tot een rode luminescentie, afkomstig van het complex Li_{int}-Li_{Ga}-O_P. De intensiteit ervan is ongeveer een factor 30 lager dan die van de COL. *b*) Bij een GaP-kristal met Zn-dotering dat op overeenkomstige wijze gekweekt is, en dat vervolgens gestookt is gedurende 50 uur bij 500 °C. De getrokken curve geeft het spectrum van het nog niet gestookte kristal. De onderbroken lijn laat zien wat het stoken teweeg brengt: de COL wordt onderdrukt en tegelijk neemt men een sterke stijging waar van de emissie in het rood, die afkomstig is van het Zn-O-complex.

1. De elementen uit kolom II van het periodieke systeem, zoals Be, Mg, Zn, en Cd, kunnen de Ga-vacature in het V_{Ga}-O_P-complex opvullen, en daardoor naaste-buurcomplexen doen ontstaan die nuttig licht emitteren (b.v. Zn_{Ga}^{-}-O_P^{+}). Dit vormingsmechanisme treedt vermoedelijk op vanwege het volgende. In geladen toestand is het V_{Ga}-O_P-complex een dubbele acceptor, $(V_{Ga}$-$O_P)^{--}$, wat volgt uit magneto-optische gegevens over de scherpe A-lijn in de COL [50]. Men

[*] Zie echter voetnoot [79].

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

veronderstelt dat het Zn-atoom in GaP diffundeert door een interstitie-substitutiemechanisme. Verder is een interstitieel Zn-atoom (enkelvoudig) positief geladen, en heeft dus een donorkarakter. De Coulomb-wisselwerking tussen Zn_{int}^{+} en $(V_{Ga}\text{-}O_P)^{--}$ kan aldus tot de vorming van een neutraal complex $Zn_{Ga}^{-}\text{-}O_P^{+}$ leiden:

$$Zn_{int}^{+} + (V_{Ga}\text{-}O_P)^{--} \rightarrow Zn_{Ga}^{-}\text{-}O_P^{+} + e. \qquad (7)$$

Men ziet dat de V_{Ga}-O_P-complexen een beslissende rol vervullen bij de vorming van de Zn_{Ga}-O_P-complexen met stralingsemissie, en daaruit volgt weer dat men voor het beheersen van de concentratie der Zn_{Ga}-O_P-complexen noodzakelijkerwijs de concentratie van de V_{Ga}-O_P-complexen moet kunnen beheersen en optimaliseren [51][54]. We komen hierop terug in paragraaf VI-C.

2. Een tweede belangrijk aspect van de V_{Ga}-O_P-complexen is hun gedrag in kristallen met Zn-dotering die gemaakt zijn met behulp van een waterdamptransportprocédé. De spectra vóór en na stoken zijn in fig. 15*b* afgebeeld. Na het stoken blijkt een sterke doving van de COL te zijn opgetreden, terwijl de emissie in het rood, bij 1,86 eV, afkomstig van het Zn-O-complex sterker geworden is; het laatste duidt op een stijging van het aantal stralende centra. Anderzijds ziet men dat de intensiteit van het groene licht, afkomstig van het met Zn en S samenhangende donor-acceptor-paarspectrum, eveneens toeneemt, wat erop wijst dat het V_{Ga}-O_P-complex bij lage temperatuur wel eens als niet-stralend centrum zou kunnen werken.

3. Ten slotte blijkt men het doven van de COL na stoken bij temperaturen van 400-500 °C, te kunnen toeschrijven aan de vorming van centra zoals C_{Ga}^{+}-O_P^{+} of Si_{Ga}^{+}-O_P^{+}, die als niet-stralende centra werken [45][53]. In hoofdstuk VIII zullen we bespreken hoe deze centra dit doen. De vorming van deze centra, b.v. voor Si, geschiedt volgens een formule die op verg. (7) lijkt, en die luidt:

$$Si_{int}^{+} + (V_{Ga}\text{-}O_P)^{--} \rightarrow Si_{Ga}^{+}\text{-}O_P^{+} + 3e. \qquad (8)$$

Op deze wijze zullen in GaP-kristallen gedurende het stoken tegelijkertijd stralende centra zoals Zn_{Ga}-O_P en niet-stralende centra zoals Si_{Ga}^{+}-O_P^{+} ontstaan. Om het quantumrendement van het materiaal zo gunstig mogelijk te maken, dienen we de reactieconstanten die de vorming van deze complexen beheersen, goed te kennen. Er wordt thans op dit gebied vrij veel werk verricht, en we zullen in paragraaf VI-C de belangrijkste resultaten bespreken die men tot dusver heeft verkregen.

Een belangrijke ontdekking die men bij de bestudering van de V_{Ga}-O_P-complexen heeft gedaan, is dat het element Mg de plaats van de Ga-vacature kan innemen, waardoor een vroeger nog niet bekende luminescentie in het geel-oranje ontstaat. Men schrijft deze emissie toe aan de recombinatie van een gebonden exciton aan een Mg-O-complex van het naaste-buurtype [52][55]. De details van de optische eigenschappen van deze emissie bespreken we in hoofdstuk VII.

VI-C. De vorming van de V_{Ga}-O_P-complexen

Om te bereiken dat er meer licht emitterende Zn-O-complexen worden gevormd, moet men blijkens (7) het aantal V_{Ga}-O_P-complexen zo groot mogelijk maken. Aangezien de COL toegeschreven wordt aan de overgang van een gebonden exciton bij een V_{Ga}-O_P-complex, zou de intensiteit het aantal V_{Ga}-O_P-complexen moeten weerspiegelen. We zullen hier in het kort de experimenten beschrijven die onlangs gedaan zijn met kristallen van GaP die men heeft laten groeien met een waterdamptransportprocédé, en waarbij het erom ging met behulp van spectraal onderzoek aan de COL het aantal V_{Ga}-O_P-complexen te maximaliseren [51].

Verschillende stukjes van op deze wijze gemaakte GaP-kristal zijn gestookt bij temperaturen, uiteenlopend van 550 tot 1000 °C, waarna een zeer snelle afkoeling plaatsvond. De stooktijd lag tussen 2 uur bij de lagere temperaturen en 15 minuten bij de hoogste. In *fig. 16* is de bij deze kristallen gevonden intensiteit van de oranje fotoluminescentie (COL) uitgezet als functie van de stooktemperatuur. Er blijkt een scherp maximum voor te komen tussen twee temperatuurgebieden, waar de activeringsenergie, dat is de vormingswarmte, respectievelijk 3,3 en 1,25 eV bedraagt. Het is interessant om op te merken dat de steile daling van de intensiteit aan weerszijden van het maximum twee zeer verschillende oorzaken heeft: in het gebied

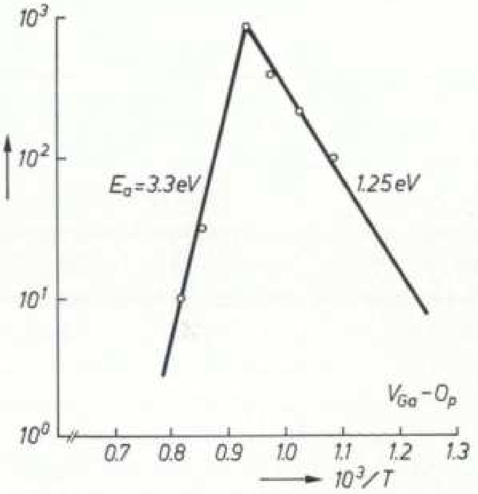


Fig. 16. Relatieve intensiteit voor het emissiespectrum van het V_{Ga}-O_P-complex, als functie van de stooktemperatuur *T*. De activeringsenergieën E_a voor de reacties die in het gebied van de lage en de hoge temperaturen optreden, zijn respectievelijk 1,25 eV en 3,3 eV.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

van de lage temperaturen worden centra gevormd op de wijze van vergelijking (8), aan de andere kant, bij de hoge temperaturen, treedt dissociatie van het complex galliumvacature/zuurstofdonor op. Inzicht in beide processen is van essentieel belang voor het kunnen vergroten van het quantumrendement bij GaP met (Zn,O)-dotering, aangezien het ene proces meer niet-stralende centra oplevert en het andere de complexen vernietigt die verantwoordelijk zijn voor het ontstaan van licht emitterende centra. Uitgaande van deze overwegingen trachten wij thans het quantumrendement van GaP met (Zn,O)-dotering te verbeteren [56].

VII. Mg-O-complexen van het naaste-buurtype

Aangezien een Mg-acceptor kleiner van afmetingen en minder sterk elektronegatief is dan Zn, mag men verwachten dat de emissie door een exciton, gebonden aan een Mg_{Ga}-O_P-complex, bij een wat hogere energie zal liggen dan het geval is bij een Zn_{Ga}-O_P-complex [52] [55]. Onlangs heeft men bij GaP met (Mg,O)-dotering inderdaad oranje-gele straling waargenomen die geëmitteerd wordt met een goed rendement en waarvan de maximum-intensiteit bij 6030 Å (bij 300 K) ligt [52] [55]. Deze emissie wordt toegeschreven aan excitonen, gebonden aan een naaste-buurcomplex [52]. De rol die de zuurstof speelt, heeft men weer kunnen bevestigen [55] door een isotopenproef met ^{18}O.

Fig. 17a toont het spectrum van deze oranje-gele luminescentie in een preparaat waarin Mg is gediffundeerd. Het spectrum bestaat uit twee aparte banden: een deel met een zekere structuur dat als spectrum *A* is aangeduid en een structuurloos deel *B*. De betrokken energieniveaus (fig. 17*b*) heeft men gevonden door metingen van de uitsterftijd (fig. 17*c*) als functie van de temperatuur en door luminescentie-excitatieproeven. De spectra *A* en *B* schrijft men toe aan de recombinatie van een elektron in een Mg-O-complex met een gat in het Mg-O-excitonniveau E_h, of met een gat in een geïsoleerd Mg-acceptorniveau E_a; de laatstgenoemde recombinatie leidt dus tot een paarspectrum (*B*). De situatie lijkt op het dubbele spectrum (van het exciton- en het paartype) dat in hoofdstuk V voor de Cd-O- en de Zn-O-complexen is besproken. De intensiteitsverhouding van de spectra *A* en *B* hangt ten nauwste af van de verhouding (Mg-O)/(Mg). Zoals ook reeds in paragraaf V-A ter sprake kwam, mag men evenwel verwachten dat bij kamertemperatuur de emissie voornamelijk een gevolg is van excitonovergangen.

Het is van belang op te merken dat bij zeer snel afkoelen vanaf 950 °C de Mg-O-complexen kunnen worden vernietigd, waarbij tegelijkertijd de COL, afkomstig van V_{Ga}-O_P-complexen, weer terugkomt. Dit is een sterke aanwijzing dat de vorming van Mg-O wordt voorafgegaan door die van een V_{Ga}-O_P-complex. Zoals eerder is opgemerkt, moet men daarom de optimalisering van de V_{Ga}-O_P-complexen beschouwen als een belangrijke stap voor het verbeteren van het rendement bij de galliumfosfide-LED's waarin iso-elektronische naaste-buurcomplexen als licht emitterende centra optreden.

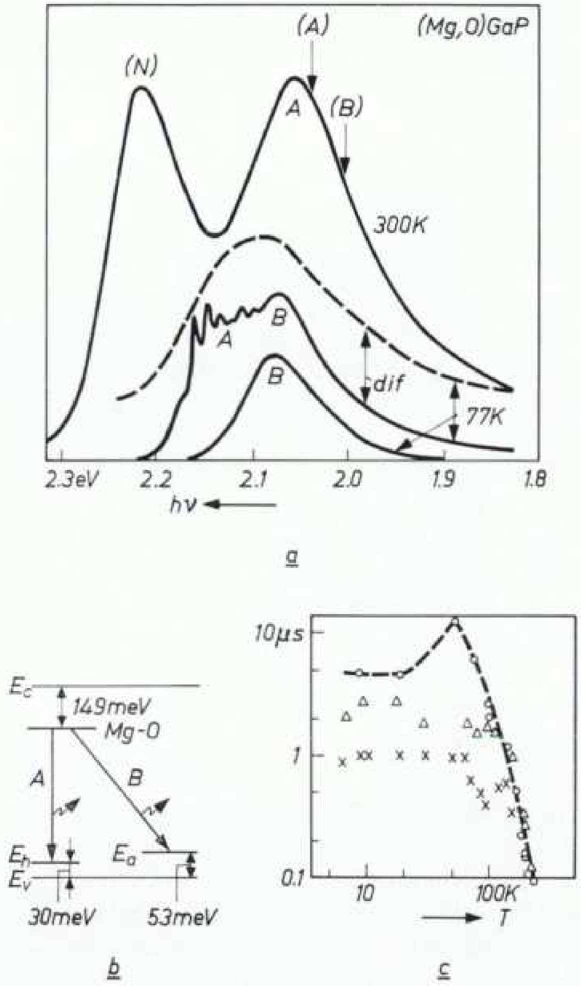


Fig. 17. De recombinatieeigenschappen van aan het Mg-O-complex gebonden elektronen. *a*) De fotoluminescentiespectra (getrokken lijnen) en elektroluminescentie (streeplijn), voor kristallen waarin Mg gediffundeerd is (*dif*) en voor kristallen die uit een oplossing verkregen zijn. Bij de preparaten met Mg-diffusie treden zowel een excitonspectrum (*A*) als een, structuurloos, paarspectrum (*B*) op; in het geval van de uit een oplossing verkregen kristallen domineert het paarspectrum *B* (bij 77 K). Bij een temperatuur van 300 K is de luminescentie afkomstig van de excitonrecombinatie (piek *A*). De verticale pijlen gemerkt (*A*) en (*B*) geven de verwachte positie van de maxima voor de beide typen emissie, in de veronderstelling dat de temperatuurafhankelijkheid van het elektronniveau van het Mg-O-complex en die van de bandafstand dezelfde zijn. De piek bij 2,22 eV is een zg. A-lijn, die te maken heeft met een overgang bij een N-atoom. *b*) Schematische voorstelling van de betrokken energieniveaus; de exciton- en paarovergangen zijn respectievelijk aangeduid met *A* en *B*. E_h is de bindingsenergie van een gat in een excitonniveau en E_a die in een acceptorniveau. *c*) De karakteristieke uitsterftijd (1/e-tijd) als functie van de temperatuur voor een kristal dat uit een oplossing verkregen is (cirkeltjes), voor een laagje dat met het LPE-procédé is gemaakt (driehoekjes), en voor een preparaat dat door diffusie is gedoteerd (kruisjes). De resultaten doen denken aan die voor GaP met een (Zn,O)-dotering, bij zwakke excitatie [43].

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Het uitwendige quantumrendement van de fotoluminescentie in GaP met (Mg,O)-dotering blijkt bij 100 K groter dan 50% te zijn. Wij hebben diodes gemaakt door een *N*-laag met Te-dotering te laten groeien op uit een oplossing verkregen kristallen. Bij deze diodes heeft men bij 77 K een uitwendig quantumrendement van ca. 0,1% gemeten. Verhoging van het rendement bij kamertemperatuur vereist een groter aantal Mg-O-complexen in de buurt van de *P-N*-overgang en/of een flinke vermindering van het aantal niet-stralende centra, die het normale vangproces beconcurreren door de vangst van elektronen die langs thermische weg uit de relatief ondiep gelegen Mg-O-complexen ontsnappen.

VIII. Niet-stralende centra

Recombinatieprocessen die stralingsloos verlopen — de energie komt dan vrij als warmte (b.v. door roostertrillingen), maar niet als licht — zijn moeilijk te onderzoeken aangezien een rechtstreekse waarneming ervan bijna onmogelijk is. Dit valt te betreuren, omdat deze processen van zeer groot belang zijn voor de luminescentie-eigenschappen van halfgeleiders met een indirecte kleinste bandafstand. Het quantumrendement in GaP bij kamertemperatuur is daar een voorbeeld van; het wordt beperkt door de recombinatie van elektronen en gaten bij deze niet-stralende centra, zoals we in V-B hebben gezien. Een bespreking van de aard van deze centra is derhalve van belang; hun concentratie zal verkleind moeten worden, wil men LED's kunnen maken met een hoger quantum-rendement.

Het belangrijkste stralingsloze proces in deze halfgeleiders is het Auger-proces. Bij het Auger-proces vindt men de energie die bij de recombinatie van een elektron en een gat vrijkomt niet terug als foton, maar als kinetische energie van een in de buurt zijnd derde deeltje (elektron of gat). Dit derde deeltje kan vrij of gebonden zijn. Verschillende gevallen zijn onderzocht en we zullen enige ervan, die in *fig. 18* apart zijn afgebeeld, bespreken.

Het eerste geval heeft betrekking op GaP met S-dotering; zwavel is hier de ondiepe donor. Bij lage temperaturen kan de neutrale zwaveldonor een exciton binden, wat aanleiding geeft tot een exiton-overgang van 2,309 eV bij 4,2 K, de C-lijn [57]. De gemeten uitsterftijd van deze luminescentie is 21 ns. Uit absorptiemetingen en elektrische metingen kan men schatten dat de uitsterftijd van bovengenoemde overgang ongeveer 11 μs moet zijn. Dit betekent een discrepantie van een factor 500 tussen de waargenomen en de geschatte uitsterftijd. Deze discrepantie wordt begrijpelijk wanneer er naast de stralingsovergang die voor de C-lijn verantwoordelijk is, nog een stralingsloos verval via een Auger-proces optreedt. In dit proces neemt het elektron dat aan een S-atoom gebonden zit, de energie op die bij de annihilatie van een exciton ontstaat (fig. 18*a*); dit stralingsloze proces is ongeveer 500 keer zo snel als het stralende.

Het is bekend dat alle neutrale acceptors [58] [59] en donors [60] bij lage temperatuur excitonen kunnen binden, en dat de resulterende recombinatieprocessen hoofdzakelijk stralingsloos verlopen. Bij kamertemperatuur, echter, treden deze centra niet als niet-stralende centra op, omdat bij die temperatuur de excitonen reeds gedissocieerd zijn.

Het tweede soort Auger-processen heeft te maken met neutrale centra van een ander type, zoals het Zn-O-complex (fig. 18*b*). In *P*-materiaal met (Zn,O)-dotering dat een goed luminescentierendement vertoont, zijn vrije gaten bij kamertemperatuur in overvloed aanwezig (ca. 5×10^{17} cm^{-3}). Steeds wanneer het gebonden gat van het exciton recombineert met het elektron (dat sterk gebonden is), bestaat de mogelijkheid de energie als kinetische energie aan een vrij gat over te dragen; dit is een Auger-proces [61] [62]. Ten einde een groot aantal Zn-O-complexen te verkrijgen, moet de Zn-concentratie natuurlijk hoog zijn [63]. Bij deze hoge concentratie blijkt de niet-stralende recombinatie door het Auger-proces evenwel te domineren [41] [42]; Zn, als acceptor, levert namelijk de vrije gaten. Er zal dus een optimale waarde voor de concentratie van het zink bestaan; de beste diodes zijn die waarbij de concentratie ongeveer 3×10^{17} cm^{-3} is [18].

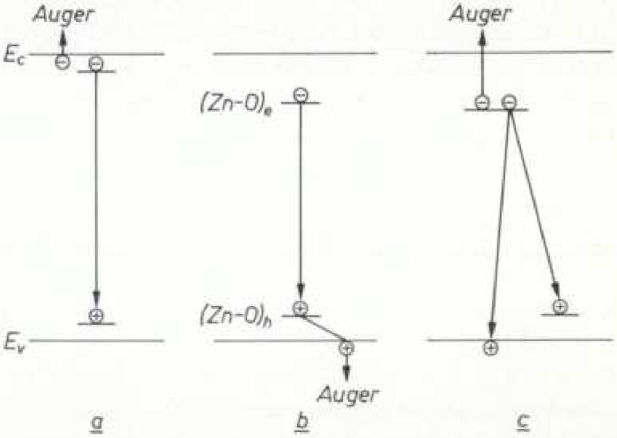


Fig. 18. Schematische voorstelling van drie soorten Auger-processen. Bij deze processen wordt de energie die vrijkomt bij de recombinatie van een elektron met een gat, als kinetische energie meegedeeld aan een naburig derde deeltje (elektron of gat). *a*) Auger-proces dat kan optreden bij een neutrale zwaveldonor, als gevolg van de annihilatie van een exciton. *b*) Auger-proces bij een Zn-O-complex, als gevolg van de annihilatie van een gebonden exciton; daarbij fungeert een vrij gat in de valentieband als derde deeltje. Een variant op deze situatie is die waarbij het elektron dat gebonden is aan het Zn-O-complex, recombineert met het *vrije* gat, en het *gebonden* gat bij het complex als derde deeltje fungeert. *c*) Auger-proces bij een tweevoudig ioniseerbare donor. Het gebonden elektron kan recombineren met een vrij of met een gebonden gat. De daarbij ontstane energie wordt overgedragen aan het tweede gebonden elektron bij de donor.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Er bestaat een andere benaderingswijze voor het bereiken van een hoger rendement; hierbij poogt men een groot aantal Zn-O-complexen aan te brengen zonder daarbij te doteren met een onnodig groot aantal Zn-atomen. Onderzoek in deze richting is gaande [51].

Het is de moeite waard ten aanzien van de twee soorten iso-elektronische centra (te weten het N-atoom en het Zn-O-complex) nog een opmerking te maken. Zoals we zagen, leiden grote concentraties van zink overwegend tot stralingsloze recombinatieprocessen. Iso-elektronische centra als N-atomen, daarentegen, voegen aan het kristal géén ladingdragers toe, en kunnen daardoor gelukkig de stralingsloze recombinatieprocessen niet bevorderen — b.v. een hier nog niet genoemd type Auger-proces, waarbij een door de sterke dotering ontstane „impurity band" betrokken is [64]. De LED's voor groen licht met het beste rendement hebben dan ook een N-dotering die groter is dan 10^{19} cm^{-3}, en de GaP-diodes voor geel licht bevatten per cm^3 ongeveer 10^{22} N-atomen [65].

Ten slotte zullen we een groep niet-stralende centra bespreken met diepe energieniveaus, die bij kamertemperatuur inderdaad als niet-stralend centrum kunnen fungeren. Deze centra zijn de dubbel geïoniseerd, en kunnen dus *twee* deeltjes van dezelfde soort binden. Nemen we b.v. het geval van GaP met Cl-dotering. Aangezien Cl een element is uit de zevende kolom van het periodieke systeem, kan het fosfor vervangen, waarbij diepe, dubbel geïoniseerde donors ontstaan [45] [66]. Normaal kunnen Cl^{++}-atomen twee elektronen vangen. Wanneer er nu ook nog vrije gaten beschikbaar zijn, kan recombinatie optreden. De daarbij vrijkomende energie vindt men evenwel niet in fotonen terug; in plaats daarvan wordt de energie overgedragen aan het tweede elektron dat door het Cl-atoom gebonden is (fig. 18*c*). Op overeenkomstige wijze zullen, wanneer toevallig aanwezige verontreinigingen zoals Si of C de galliumvacature in een V_{Ga}-O_P-complex opvullen, de dubbel geïoniseerde complexen $Si_{Ga}{}^+$-$O_P{}^+$ en $C_{Ga}{}^+$-$O_P{}^+$ ontstaan. Deze centra zijn eveneens tweevoudig geladen en zullen Auger-recombinatieprocessen sterk bevorderen [53] [67]. Twee eigenschappen van deze dubbel geïoniseerde centra zijn de volgende:

(a) Aangezien de centra dubbel geïoniseerd zijn, mag men verwachten dat de werkzame doorsnede voor vangst groot zal zijn [67], wat impliceert dat zij de recombinatiekinetica in GaP zullen kunnen beheersen.

(b) Verontreinigingen als Cl en Si zijn al ongewild aanwezig in de meeste GaP-kristallen, en in feite kunnen deze verontreinigingen voor de huidige beperking van het quantumrendement een belangrijke factor zijn. Verwijderen ervan zou tot een beter rendement moeten leiden, zowel in de diodes als bij fotoluminescentie.

Hieraan kan nog een interessante opmerking worden toegevoegd over de verschillende groeiprocessen die men gebruikt bij de vervaardiging van GaP-diodes. Onlangs heeft men LED's gemaakt, zowel volgens de VPE-methode („vapour-phase epitaxy" [68]) als de LPE-methode („liquid-phase epitaxy" [69]), waarvan de groene luminescentie wordt toegeschreven aan de recombinatie van vrije excitonen met fononassistentie. In deze diodes lijken het quantumrendement, dat niet groot is, en de uitsterftijd van de luminescentie begrensd te worden door de aanwezigheid van niet-stralende centra. In beide typen diode heeft men dezelfde uitsterftijd (400 ns) waargenomen, wat mogelijkerwijs betekent dat de concentratie van niet-stralende verontreinigingen bij VPE en LPE vrijwel gelijk is.

IX. Slotopmerkingen en verwachte ontwikkelingen

Uit de gegeven bespreking van licht emitterende centra mag men concluderen dat de recente verbeteringen van quantumrendement en kleurenscala voor LED's van GaP een veel belovende toekomst inhouden. Er blijft echter nog één vraag te beantwoorden: hoe hoog ligt de theoretische bovengrens voor het uitwendige quantumrendement bij deze halfgeleider? Vertaald in praktische termen: wat is de kans dat we ooit een vastestoflamp zullen kunnen maken die een lichtstroom in de grootte van lumens opwekt?

Tot dusver zijn geen theoretische voorspellingen gedaan die gebaseerd zijn op de fysische verschijnselen die zich kunnen voordoen bij recombinatie ter plaatse van gelokaliseerde centra in GaP. Vanwege het belang van de gestelde vraag zullen we een semi-kwantitatief antwoord [70] pogen te geven in het geval van recombinatie bij het Zn-O-complex, eigenlijk het voornaamste centrum dat in dit artikel besproken is.

Ten einde een hoger quantumrendement in GaP met (Zn,O)-dotering te verwezenlijken, dient men de verhouding tussen het aantal stralende centra (in dit geval het aantal Zn-O-complexen) en het aantal niet-stralende centra te vergroten. De vermindering van het aantal niet-stralende centra kan men bereiken door (a) de Zn- concentratie zo klein mogelijk te houden, waardoor men het aantal Auger-processen met vrije gaten aan de *P*-zijde vermindert, en door (b) de niet-stralende centra zoals C^+-O^+ en Si^+-O^+, enz. te verminderen.

Tegenwoordig behaalt men ca. 3×10^{16} Zn-O-complexen per cm^3, bij een Zn-concentratie van 3×10^{17} per cm^3, of, in formule geschreven, $N_{Zn-O}/N_{Zn} \approx 0{,}1$. Door de thermodynamica van de vorming van het complex beter te bestuderen, mag men verwachten dat $N_{Zn-O}/N_{Zn} \approx 1$ haalbaar zal worden. De eerste berekeningen hierover [71] wijzen erop dat men om N_{Zn-O}/N_{Zn} te maximaliseren, het zink pas moet inbou-

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

wen *nadat* het aantal V_{Ga}-O_P-complexen gemaximaliseerd is. Verder kan men niet-stralende centra als Si_{Ga}^+-O_P^+ of Cl^{++} vermijden door kristallen te maken waarin deze verontreinigingen niet voorkomen. Als alternatief bestaat dan nog de mogelijkheid van een geschikte warmtebehandeling, die de concentratie van verontreinigingen als Si_{Ga}^+-O_P^+ kan minimaliseren [56].

Veronderstellen we nu dat inderdaad $N_{Zn-O}/N_{Zn} = 1$ is en dat door een geschikte wijze van groeien en een passende warmtebehandeling de concentratie van de niet-stralende centra aanzienlijk verminderd is. Volgens het in V-B besproken model zal de veronderstelde situatie corresponderen met $\tau_n/\tau_{nt} \gg 1$, en met behulp van de geschatte variatie van τ_D, τ_R en τ_{NR} als functie van de concentratie der vrije gaten (fig. 12), mogen we uit dit alles voorspellen dat $\tau_R \approx 3000$ ns en $\tau_D \approx 1500$ ns voor $N_{Zn} \approx 10^{17}$ cm^{-3}. Uit verg. (5) volgt dan dat η ca. 50% bedraagt. Zulk een hoog quantumrendement is wellicht in de naaste toekomst al bereikbaar.

Bij quantumrendementen van 50% voor rode LED's kan men als toepassing gaan denken aan vlakke lichtbronnen voor indicators met extra grote oppervlakte. Laten we de situatie eens vergelijken met een gloeilamp van 15 W, waarbij ongeveer 10% van het vermogen (of 1,5 W) in zichtbaar licht en 1% (of 0,15 W) in rood licht wordt omgezet. Het lichtstroomrendement van een dergelijke lamp is ongeveer 10 lm/W, hetgeen overeenkomt met ongeveer 1 lm/W in het rode deel van het zichtbare spectrum. Een rode LED met een quantumrendement van 50% zou bij 2 volt gelijkspanning en een stroom van 50 mA, 0,05 W aan rood licht opwekken. Een combinatie van 3 LED's levert dan een optisch vermogen van 0,15 W, de waarde die ook voor de vermelde gloeilamp in het rode gebied geldt. Verder correspondeert het lichtstroomrendement van deze rode LED's door het hogere quantumrendement met 10 lm/W (men zie ook tabel I).

Aan de stralende centra die in GaP de emissie in het groen en in het geel opwekken, moet nog veel meer onderzocht worden voor men tot dergelijke beschouwingen kan komen. Aangezien het oog in die kleurgebieden echter veel gevoeliger is, zouden quantumrendementen van 1 tot 10% toch reeds overeenkomen met 5-50 lm/W.

Onlangs heeft men geopperd dat het hoge rendement in GaP met N- of (Zn,O)-dotering wel eens zou kunnen leiden tot gestimuleerde emissie van coherente straling, met andere woorden tot laserwerking, en er zijn daarover experimenten gedaan [72] [73]. In plaats van de verwachte lijnversmalling heeft men een karakteristieke inhomogene verbreding waargenomen [72]; verder werk is nodig om na te gaan of er in het zichtbare deel van het spectrum bij GaP laserwerking optreedt. Naast GaP en de andere ternaire III-V-verbindingen zijn er nog enige halfgeleiders die wellicht in aanmerking komen voor onderzoek aan stralende centra. Een voorbeeld is SiC, dat blauw licht kan emitteren; bij injectie treedt luminescentie op, zowel bij *N*- als bij *P*-materiaal. In kristallografische termen is SiC een materiaal met polytypie; van de vele modificaties zijn voor de luminescentie het interessantst het kubische type (β-SiC) en een hexagonaal type (α-SiC met 6H-structuur), die bij 0 K een indirecte kleinste bandafstand hebben van 2,38 eV, respectievelijk 3,08 eV. Aangezien het smeltpunt van SiC ongeveer 2600 °C bedraagt, is het maken van eenkristallen van deze halfgeleider een moeilijke zaak. Daarbij komt nog dat de fysische en de chemische eigenschappen van de niet-stralende recombinatiecentra in SiC niet grondig bestudeerd zijn, hoofdzakelijk omdat tot dusver geen materiaal met een goed beheerste dotering gemaakt is. Wel heeft men epitactische lagen laten groeien, waardoor LED's ontstaan die in het geel en in het blauw kunnen emitteren; de lichtopbrengst is betrekkelijk slecht, althans bij een injectie van ladingdragers op een niveau dat nog aanvaardbaar genoemd mag worden [74] (tabel I).

Dan is er nog GaN. Bij deze halfgeleider, die een verticale kleinste bandafstand van 3,50 eV heeft, zijn onlangs hoge rendementen voor fotoluminescentie gevonden [75] (12% bij 300 K b.v.) en laserwerking door optisch pompen [76]. Het fundamentele probleem bij deze halfgeleider ligt, behalve in het verkrijgen van een goede methode voor het maken van eenkristallen, in het feit dat tot dusver alleen maar GaN van het type *N* is gemaakt. Het maken van *P*-GaN is wellicht niet goed mogelijk, zoals men ook heeft ervaren bij andere halfgeleiders met een grote verticale bandafstand, zoals de II-VI-verbindingen. Tot heden heeft men alleen GaN-diodes vervaardigd die bestaan uit de overgang tussen isolerend materiaal en *N*-materiaal verkregen door Zn-diffusie. Bij kamertemperatuur vertonen deze een gelijkstroomluminescentie in het groen en in het blauw [77]. Het uitwendige vermogensrendement is 10^{-4} en er is een uitwendig quantumrendement van 1% behaald.

Hiermee zijn we aan het einde van een overzicht van de voortgang die het onderzoek aan stralende centra en niet-stralende centra in GaP heeft gemaakt. Het is hierin duidelijk gebleken hoe gunstig de fysische en chemische kanten van dit onderzoek op elkaar inwerken. Hoe ver men met de LED's nog zal komen op de weg naar echte vastestoflampen hangt er op het ogenblik voornamelijk van af of we in staat zijn de nieuw verkregen kennis te vertalen in diodes met een beter rendement, onder gebruikmaking van commercieel reeds aantrekkelijk gebleken procédés. Het lijkt erop dat de vastestoflampen een goede toekomst tegemoet gaan.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

De schrijver is dank verschuldigd aan dr. D. D. King en aan dr. S. K. Kurtz voor hun waardevol commentaar en hun steun, en tevens aan ir. A. T. Vink (Eindhoven) voor nuttige opmerkingen.

Literatuur

[1] O. W. Lossew, Telegrafia i Telefonia **18**, 61, 1923, en C.R. Acad. Sci. U.S.S.R. **29**, 360 en 363, 1940.
[2] G. Destriau, J. Chimie phys. **33**, 587, 1936.
[3] A. G. Fischer, Electroluminescence in II-VI compounds, in: P. Goldberg (redacteur), Luminescence of inorganic solids, Academic Press, New York 1966, blz. 541-602.
[4] M. I. Nathan, W. P. Dumke, G. Burns, F. H. Dill Jr. en G. Lasher, Appl. Phys. Letters **1**, 62, 1962.
R. N. Hall, G. E. Fenner, J. D. Kingsley, T. J. Soltys en R. O. Carlson, Phys. Rev. Letters **9**, 366, 1962.
T. M. Quist, R. H. Rediker, R. J. Keyes, W. E. Krag, B. Lax, A. L. McWhorter en H. J. Zeiger, Appl. Phys. Letters **1**, 91, 1962.
[5] J. W. Allen, M. E. Moncaster en J. Starkiewicz, Solid-State Electronics **6**, 95, 1963.
[6] H. G. Grimmeiss en H. Scholz, Physics Letters **8**, 233, 1964.
[7] M. Gershenzon, Electroluminescence from p-n junctions in semiconductors, in: P. Goldberg (redacteur), Luminescence of inorganic solids, Academic Press, New York 1966, blz. 603-684.
M. Gershenzon, Radiative recombination in the III-V compounds, in: R. K. Willardson en A. C. Beer (redacteuren). Semiconductors and semimetals Vol. 2, Academic Press, New York 1966, blz. 289-369.
[8] D. G. Thomas, Physics Today **21**, no. 2, 43, feb. 1968.
[9] M. R. Lorenz, Trans. Met. Soc. AIME **245**, 539, 1969.
[10] D. G. Thomas, Brit. J. appl. Phys./J. Phys. D **2**, 637, 1969.
[11] Y. P. Varshni, Phys. Stat. sol. **19**, 459, 1967, en **20**, 9, 1967.
[12] P. J. Dean, Junction electroluminescence, in: R. Wolfe en C. J. Kriessman (redacteuren), Applied solid state science Vol. 1, Academic Press, New York 1969, blz. 1-151.
[13] P. J. Dean, J. Luminescence **1/2**, 398, 1970.
[14] H. C. Casey Jr. en F. A. Trumbore, Mat. Sci. Engng. **6**, 69, 1970.
[15] D. G. Thomas, IEEE Trans. **ED-18**, 621, 1971.
[16] A. A. Bergh en P. J. Dean, Proc. IEEE **60**, 156, 1972.
[17] I. Ladany, J. appl. Phys. **42**, 654, 1971.
[18] Solomon c.s. hebben onlangs door een dompelprocédé GaP-diodes kunnen maken die rood licht uitzenden met een uitwendig quantumrendement van 12% (zie R. Solomon en D. DeFevere, J. Electronics Mat. **1**, 26, 1972). Hun techniek maakt gebruik van het tweevoudige epitactische groeiprocédé beschreven door R. H. Saul, J. Armstrong en W. H. Hackett Jr., Appl. Phys. Letters **15**, 229, 1969, waarbij uitwendige quantumrendementen van 7,2% werden behaald.
[19] I. Hayashi, M. B. Panish, P. W. Foy en S. Sumski, Appl. Phys. Letters **17**, 109, 1970.
[20] F. W. Ostermayer Jr., Metall. Trans. **2**, 747, 1971.
S. V. Galginaitis, Metall. Trans. **2**, 757, 1971.
[21] J. J. Hopfield, D. G. Thomas en M. Gershenzon, Phys. Rev. Letters **10**, 162, 1963.
[22] P. J. Dean, C. H. Henry en C. J. Frosch, Phys. Rev. **168**, 812, 1968.
[23] R. N. Bhargava, Phys. Rev. B **2**, 387, 1970; zie ook J. M. Dishman, Phys. Rev. B **3**, 2588, 1971.
[24] P. J. Dean en C. H. Henry, Phys. Rev. **176**, 928, 1968.
[25] H. Kukimoto, C. H. Henry en G. L. Miller, Bull. Amer. Phys. Soc. **17**, 233, 1972.
[26] D. G. Thomas, J. J. Hopfield en C. J. Frosch, Phys. Rev. Letters **15**, 857, 1965.
[27] R. A. Logan, H. G. White en W. Wiegmann, Solid-State Electronics **14**, 55, 1971.
[28] J. J. Hopfield, D. G. Thomas en R. T. Lynch, Phys. Rev. Letters **17**, 312, 1966.
[29] R. A. Faulkner, Phys. Rev. **175**, 991, 1968.
[30] J. C. Phillips, Phys. Rev. Letters **22**, 285, 1969.
[31] T. N. Morgan, B. Welber en R. N. Bhargava, Phys. Rev. **166**, 751, 1968.
[32] C. H. Henry, P. J. Dean en J. D. Cuthbert, Phys. Rev. **166**, 754, 1968.
[33] M. Gershenzon en R. M. Mikulyak, Appl. Phys. Letters **8**, 245, 1966.
[34] K. Maeda, M. Naito en A. Kasami, Jap. J. appl. Phys. **8**, 817, 1969.
[35] M. R. Lorenz en M. Pilkuhn, J. appl. Phys. **37**, 4094, 1966.
[36] H. Nelson, RCA Rev. **24**, 603, 1963.
[37] S. J. Bass en P. E. Oliver, J. Crystal Growth **3/4**, 286, 1968.
[38] M. Gershenzon, F. A. Trumbore, R. M. Mikulyak en M. Kowalchik, J. appl. Phys. **36**, 1528, 1965.
[39] J. D. Cuthbert, C. H. Henry en P. J. Dean, Phys. Rev. **170**, 739, 1968.
[40] A. Onton en M. R. Lorenz, Appl. Phys. Letters **12**, 115, 1968.
[41] R. N. Bhargava, J. appl. Phys. **41**, 3698, 1970.
J. S. Jayson, R. N. Bhargava en R. W. Dixon, J. appl. Phys. **41**, 4972, 1970.
[42] J. M. Dishman en M. DiDomenico Jr., Phys. Rev. B **1**, 3381, 1970.
J. M. Dishman, M. DiDomenico Jr. en R. Caruso, Phys. Rev. B **2**, 1988, 1970.
[43] J. S. Jayson en R. W. Dixon, J. appl. Phys. **42**, 774, 1971.
J. S. Jayson en R. Z. Bachrach, Phys. Rev. B **4**, 477, 1971.
[44] A. Kasami, Jap. J. appl. Phys. **9**, 946, 1970.
[45] R. N. Bhargava, Proc. Tenth Int. Conf. Phys. Semicond., Cambridge, Mass., 1970, blz. 640.
[46] M. A. Afromowitz en M. DiDomenico Jr., J. appl. Phys. **42**, 3205, 1971.
[47] J. A. W. van der Does de Bye en A. T. Vink, J. Luminescence **5**, 108, 1972.
[48] P. J. Dean, Phys. Rev. B **4**, 2596, 1971.
[49] R. N. Bhargava, S. K. Kurtz, A. T. Vink en R. C. Peters, Phys. Rev. Letters **27**, 183, 1971.
[50] P. J. Dean, Solid State Comm. **9**, 2211, 1971.
[51] G. M. Blom en R. N. Bhargava, voordracht 2nd Int. Conf. on Vapour Growth and Epitaxy, Jerusalem 1972.
[52] R. N. Bhargava, C. Michel, W. L. Lupatkin, R. L. Bronnes en S. K. Kurtz, Appl. Phys. Letters **20**, 227, 1972.
[53] R. Z. Bachrach, O. G. Lorimor, L. R. Dawson en K. B. Wolfstirn, Bull. Amer. Phys. Soc. **16**, 436, 1971.
[54] M. Toyama en A. Kasami, verschijnt in Jap. J. appl. Phys.
[55] P. J. Dean en M. Ilegems, J. Luminescence **4**, 201, 1971.
[56] R. N. Bhargava, niet gepubliceerd.
[57] D. G. Thomas, M. Gershenzon en J. J. Hopfield, Phys. Rev. **131**, 2397, 1963.
D. F. Nelson, J. D. Cuthbert, P. J. Dean en D. G. Thomas, Phys. Rev. Letters **17**, 1262, 1966.
[58] P. J. Dean, R. A. Faulkner, S. Kimura en M. Ilegems, Phys. Rev. B **4**, 1926, 1971.
[59] A. T. Vink en R. C. Peters, J. Luminescence **3**, 209, 1970.
[60] P. J. Dean, Phys. Rev. **157**, 655, 1967.
P. J. Dean, R. A. Faulkner en S. Kimura, Phys. Rev. B **2**, 4062, 1970.
A. T. Vink, A. J. Bosman, J. A. W. van der Does de Bye en R. C. Peters, J. Luminescence **5**, 57, 1972.
[61] B. Welber en T. N. Morgan, Phys. Rev. **170**, 767, 1968.
[62] K. P. Sinha en M. DiDomenico Jr., Phys. Rev. B **1**, 2623, 1970.
[63] J. C. Tsang, P. J. Dean en P. T. Landsberg, Phys. Rev. **173**, 814, 1968.
[64] J. D. Wiley, J. Phys. Chem. Solids **32**, 2053, 1971.
[65] R. Nicklin, C. D. Mobsby, G. Lidgard en P. B. Hart, J. Physics C **4**, L 344, 1971.
[66] R. N. Bhargava, Bull. Amer. Phys. Soc. **16**, 409, 1971.
[67] G. F. Neumark en D. Polder, Bull. Amer. Phys. Soc. **17**, 238, 1972, en persoonlijke mededeling.
[68] W. O. Groves, A. H. Herzog en M. G. Craford, Appl. Phys. Letters **19**, 184, 1971.
M. G. Craford, persoonlijke mededeling.
[69] R. Z. Bachrach en O. G. Lorimor, J. appl. Phys. **43**, 500, 1972.
[70] Onlangs is een schatting gedaan (J. M. Dishman, Int. Electron Devices Meeting 1970, Washington, D.C., blz. 154, Abs. 24.2), waaruit zou volgen dat een verkleining van het aantal niet-stralende centra met een factor 10 het quantumrendement van de diodes slechts een factor 2 kan doen toenemen. Dergelijke berekeningen berusten echter op een globale kennis van diverse parameters (concentratie van de stralende centra, werkzame doorsneden voor vangst, etc.), en men mag verwachten dat de resultaten aanzienlijk veranderen als deze parameters beter bekend worden.
[71] G. M. Blom, niet gepubliceerd.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

[72] N. Holonyak Jr., D. R. Scifres, H. M. Macksey, R. D. Dupuis, Y. S. Moroz, C. B. Duke, G. G. Kleiman en F. V. Williams, Phys. Rev. Letters **28**, 230, 1972.
[73] R. D. Burnham, N. Holonyak Jr., D. L. Keune en D. R. Scifres, Appl. Phys. Letters **18**, 160, 1971.
R. E. Nahory, K. L. Shaklee, R. F. Leheny en R. A. Logan Phys. Rev. Letters **27**, 1647, 1971.
[74] E. E. Violin, A. A. Kal'nin, V. V. Pasynkov, Yu. M. Tairov en D. A. Yas'kov, in: Silicon Carbide 1968, Proc. int. Conf., University Park, Penn. (special nummer van Mat. Res. Bull. **4**, 1969), blz. S 231.
R. W. Brander en R. P. Sutton, Brit. J. appl. Phys./J. Phys. D **2**, 309, 1969.
G. Kamath, persoonlijke mededeling.
[75] M. Ilegems, R. Dingle en R. A. Logan, Bull. Amer. Phys. Soc. **17**, 233, 1972.
[76] R. Dingle, K. L. Shaklee, R. F. Leheny en R. B. Zetterstrom, Appl. Phys. Letters **19**, 5, 1971.
[77] J. I. Pankove, E. A. Miller en J. E. Berkeyheiser, RCA Rev. **32**, 383, 1971.
[78] D. G. Thomas, J. J. Hopfield en W. M. Augustyniak, Phys. Rev. **140**, A 202, 1965.
J. A. W. van der Does de Bye, A. T. Vink, A. J. Bosman en R. C. Peters, J. Luminescence **3**, 185, 1970.
[79] *Toegevoegd in de drukproef:* Recente experimenten van D. L. Wight c.s. (XI. International Conference on the Physics of Semiconductors, Warschau 1972) hebben aangetoond dat COL in GaP de aanwezigheid van koper vereist. Bovendien hebben deze onderzoekers, alsmede A. T. Vink en R. C. Peters van Philips' Natuurkundig Laboratorium, aangetoond dat er geen ^{16}O-^{18}O-isotopeneffect optreedt, zodat zuurstof vermoedelijk géén rol speelt.
Moge dus de opvatting over het ontstaan van COL inmiddels veranderd zijn, de in dit artikel beschreven gedachten over de rol die het complex V_{Ga}-O_P speelt bij het totstandkomen van centra, blijven geldig.

Digitized in Heiloo the Netherlands Not for sale!

www.freeservicemanuals.info 11/14/2023

Jaargang 32, 1971, no. 10/11/12 blz. 293-440 Verschenen 19 december 1972

Digitized in Heiloo the Netherlands Not for sale!